DOCUMENT RESUME

ED 155 273 UD 018 304

AUTHOR	Carriere, Ronald A.; And Others
TITLE	Emergency School Aid Act (ESAA) Evaluation: Results of Supplemental Analyses Conducted in the Contract Extension Period.
INSTITUTION	System Development Corp., Santa Monica, Calif.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
PUB DATE	Mar 78
CONTRACT	OEC-0-73-0831; OEC-0-73-6336
NOTE	248p.; For related documents, see ED 117 229, ED 133 361-362, and UD 018 302 and 304-306
EDRS PRICE	MF-$0.83 HC-$12.71 Plus Postage.
DESCRIPTORS	*Academic Achievement; Achievement Tests; *Compensatory Education Programs; *Data Analysis; Elementary Secondary Education; *Program Evaluation; *Standardized Tests
IDENTIFIERS	*Emergency School Aid Act

ABSTRACT

This report focuses on a set of supplemental analyses that were performed on portions of the Emergency School Aid Act (ESAA) evaluation data. The goal of these analyses was to explore additional relationships in the data that might help to inform program policy, to confirm and/or further explicate some of the findings reported earlier, and to put some of the study results (particularly those concerned with student achievement) in a form that might be useful to individuals directly concerned with the operation or evaluation of educational intervention programs such as ESAA. Additional analyses of data from the overall ESAA evaluation concerned standardized achievement test data. These activities included some new analyses, but consisted primarily of efforts to reformat and document the test results in a manner that would make them more informative and useful to individuals operating or evaluating ESAA type programs. Efforts were also made to derive new data indices and composite variables from the existing data. Analyses of indepth study data were also made. These included: (1) relationships among resource use, program coordination, and school success, (2) development of causal models relating school and classroom factors to student achievement gains, and (3) effect on in depth study results of using different definitions of school success. (Author/AM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED155273

MAY 25 1978

EMERGENCY SCHOOL AID ACT (ESAA) EVALUATION: RESULTS OF SUPPLEMENTAL ANALYSES IN THE CONTRACT EXTENSION PERIOD

Ronald A. Carriere
John E. Coulson
Fred E. Cromer
William J. Doherty
Sally D. Hanes
Anne H. MacQueen
Clarence Bradford (Consultant)

March 1978

The work reported here was performed as part of the Emergency School Aid Act National Evaluation United States Office of Education Contract Numbers OEC-0-73-0831 and OEC-0-73-6336

System Development Corporation

TM-5236/025/00

UDO 18304

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

NOTE

This report covers work performed during the ESAA Evaluation contract extension period, September 1977 through March 1978. During this period, John E. Coulson was the Project Director; Fred E. Cromer was Manager for Data Analysis; Ronald A. Carriere, William J. Doherty, Sally D. Hanes, and Anne H. MacQueen were the principal analysts; Eric Nelson-Lund was the project programmer; Suzie R. Navarrette was the data processing technician; Frank Tierney provided editorial support in report preparation; and Ethel Rothstein was project secretary. Clarence Bradford served as a consultant and participated in the data analyses.

We wish to thank Dr. David Shoemaker, Project Officer for the U.S. Office of Education, for his support, guidance, and cooperation in this study.

AUTHORSHIP OF SUBSECTIONS

Authorship of the various sections of this volume is as follows:

I. Introduction: John Coulson

II. Overview of ESAA and of the Evaluation: John Coulson

III. Additional Analyses of Data From Overall Evaluation

- A. Activities to Increase Interpretability and Utility of ESAA Test Data
  1. Restandardization of ESAA Achievement Test Data: William Doherty
  2. The Distribution of Achievement Scores in a Disadvantaged Population: Fred Cromer
  3. Comparisons of Achievement Growth Obtained From Cross-Sectional and Longitudinal Samples in the ESAA Evaluation: Sally Hanes
- B. Reanalysis of Data Using Derived Indices and Composite Variables
  1. Derivation of New Composite Variables Characterizing Programs and Students: Sally Hanes and Fred Cromer
  2. Derivation of Alternative Outcome (Achievement) Measures: Clarence Bradford
  3. Reanalysis of Data Related to Program Impact: Fred Cromer
  4. Reanalyses of Data on Factors Influencing Achievement Gains (Relational Analyses): Clarence Bradford

IV. Additional Analyses of In-Depth Study Data

- A. Relationships Among Resource-Use, Program Coordination, and School Success: Anne MacQueen
- B. Development of Causal Models Relating School and Classroom Factors to Student Achievement Gains: Ronald Carriere
- C. Effect on In-Depth Study Results of Using Different Definitions of School Success: Anne MacQueen

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. OVERVIEW OF ESAA AND OF THE EVALUATION	1
III. ADDITIONAL ANALYSES OF DATA FROM OVERALL EVALUATION	3
A. ACTIVITIES TO INCREASE INTERPRETABILITY AND UTILITY OF ESAA TEST DATA	4
1. RESTANDARDIZATION OF ESAA ACHIEVEMENT TEST DATA	4
2. THE DISTRIBUTION OF ACHIEVEMENT SCORES IN A DISADVANTAGED POPULATION	38
3. COMPARISONS OF ACHIEVEMENT GROWTH OBTAINED FROM CROSS-SECTIONAL AND LONGITUDINAL SAMPLES IN THE ESAA EVALUATION	115
B. REANALYSIS OF DATA USING DERIVED INDICES AND COMPOSITE VARIABLES	127
1. DERIVATION OF NEW COMPOSITE VARIABLES CHARACTERIZING PROGRAMS AND STUDENTS	127
2. DERIVATION OF ALTERNATIVE OUTCOME (ACHIEVEMENT) MEASURES	136
3. REANALYSIS OF DATA RELATED TO PROGRAM IMPACT	142
4. REANALYSES OF DATA ON FACTORS INFLUENCING ACHIEVEMENT GAINS (RELATIONAL ANALYSES)	154
IV. ADDITIONAL ANALYSES OF IN-DEPTH STUDY DATA	159
A. RELATIONSHIPS AMONG RESOURCE USE, PROGRAM COORDINATION, AND SCHOOL SUCCESS	160
1. OVERVIEW OF THE 1975-76 IN-DEPTH STUDY	161
2. METHODOLOGY FOR REANALYSES	165
3. RESULTS	167
4. SUMMARY AND DISCUSSION	179
B. DEVELOPMENT OF CAUSAL MODELS RELATING SCHOOL AND CLASSROOM FACTORS TO STUDENT ACHIEVEMENT GAINS	189
1. OBSERVER RELIABILITY	189
2. THE RELATIONSHIP OF SCHOOL AND CLASSROOM FACTORS TO STUDENT ACHIEVEMENT GAINS	199
3. THE RELATIONSHIP OF SCHOOL TO CLASSROOM FACTORS	202
4. OTHER FINDINGS	208
C. EFFECT ON IN-DEPTH STUDY RESULTS OF USING DIFFERENT DEFINITIONS OF SCHOOL SUCCESS	209

TABLE OF CONTENTS (Cont'd)

		Page
1.	OUTCOME MEASURES AND THE TWO MAJOR COMPETING HYPOTHESES . . .	210
2.	ASSUMPTIONS AND LIMITATIONS OF THE ANALYSES	212
3.	REVIEW OF LITERATURE ON RELATIONSHIPS BETWEEN TEACHER MORALE AND STUDENT ACHIEVEMENT	214
4.	COMPARISON OF RELATIONSHIPS AMONG TEACHER MORALE AND ALTERNATE OUTCOME MEASURES IN THE ESAA IN-DEPTH STUDIES . . .	218
CORPORATE REFERENCES		228
GENERAL REFERENCES		230

LIST OF TABLES

		Page
Table 1.	Sample Changes Over Years	3
Table 2.	Mean Achievement Scores for Students Who Were in the ESAA Evaluation Sample for Two Years	8
Table 3.	Number of Items in CAT Subtests	22
Table 4.	Raw Score Performance of ESAA Third-Year Students	23
Table 5.	Scale Means and Standard Deviations	25
Table 6.	Raw Score, CTB Scale Score, and ESAA Scale Score Performance for Constant Percentile Performance Across Three Grades: National Norms	27
Table 7.	Raw Score, CTB Scale Score, and ESAA Scale Score Performance for Constant Percentile Performance Across Three Grades: ESAA Norms	28
Table 8.	CTB Scale Score and ESAA Scale Score Equivalents of Mean Reading Performance of Longitudinal Students in the Second-Year Evaluation	30
Table 9.	Raw Score, CTB Scale Score, and ESAA Scale Score Performance (Form B) for Constant Percentile Performance Across Three Grades: National Norms	34
Table 10.	Raw Score, CTB Scale Score, and ESAA Scale Score Performance (Form B) for Constant Percentile Performance Across Three Grades: ESAA Norms	35
Table 11.	Fall, Total Reading Scores: Grade 3, Test Level-2	75
Table 12.	Fall, Total Reading Scores: Grade 4, Test Level-2	76
Table 13.	Fall, Total Reading Scores: Grade 4, Test Level-3	77
Table 14.	Fall, Total Reading Scores: Grade 5, Test Level-3	78
Table 15.	Fall, Total Math Scores: Grade 3, Test Level-2	79
Table 16.	Fall, Total Math Scores: Grade 4, Test Level-2	80
Table 17.	Fall, Total Math Scores: Grade 4, Test Level-3	81
Table 18.	Fall, Total Math Scores: Grade 5, Test Level-3	82
Table 19.	Spring, Total Reading Scores: Grade 3, Test Level-2	83
Table 20.	Spring, Total Reading Scores: Grade 4, Test Level-3	84
Table 21.	Spring, Total Reading Scores: Grade 5, Test Level-3	85
Table 22.	Spring, Total Math Scores: Grade 3, Test Level-2	86
Table 23.	Spring, Total Math Scores: Grade 4, Test Level-3	87
Table 24.	Spring, Total Math Scores: Grade 5, Test Level-3	88

LIST OF TABLES (Cont'd)

		Page
Table 25.	Fall, Total Reading Scores: Grade 10, Test Level-2	89
Table 26.	Fall, Total Reading Scores: Grade 11, Test Level-2	90
Table 27.	Fall, Total Reading Scores: Grade 12, Test Level-2	91
Table 28.	Fall, Total Math Scores: Grade 10, Test Level-5	92
Table 29.	Fall, Total Math Scores: Grade 11, Test Level-5	93
Table 30.	Fall, Total Math Scores: Grade 12, Test Level-5	94
Table 31.	Spring, Total Reading Scores: Grade 10, Test Level-2	95
Table 32.	Spring, Total Reading Scores: Grade 11, Test Level-2	96
Table 33.	Spring, Total Reading Scores: Grade 12, Test Level-2	97
Table 34.	Spring, Total Math Scores: Grade 10, Test Level-5	98
Table 35.	Spring, Total Math Scores: Grade 11, Text Level-5	99
Table 36.	Spring, Total Math Scores: Grade 12, Test Level-5	100
Table 37.	Total Reading Scores: Pilot, Grade 3, Test Level-2	101
Table 38.	Total Reading Scores: Pilot, Grade 4, Test Level-2	101
Table 39.	Total Reading Scores: Pilot, Grade 4, Test Level-3	102
Table 40.	Total Reading Scores: Pilot, Grade 5, Test Level-3	102
Table 41.	Total Math Scores: Pilot, Grade 3, Test Level-2	103
Table 42.	Total Math Scores: Pilot, Grade 4, Test Level-2	103
Table 43.	Total Math Scores: Pilot, Grade 4- Test Level-3	104
Table 44.	Total Math Scores: Pilot, Grade 5, Test Level-3	104
Table 45.	Total Reading Scores: Basic, Grade 3, Test Level-2	105
Table 46.	Total Reading Scores: Basic, Grade 4, Test Level-2	105
Table 47.	Total Reading Scores: Basic, Grade 4, Test Level-3	106
Table 48.	Total Reading Scores: Basic, Grade 5, Test Level-3	106
Table 49.	Total Math Scores: Basic, Grade 3, Test Level-2	107
Table 50.	Total Math Scores: Basic, Grade 4, Test Level-2	107
Table 51.	Total Math Scores: Basic, Grade 4, Test Level-3	108
Table 52.	Total Math Scores: Basic, Grade 5, Test Level-3	108
Table 53.	Distribution of Gain Scores Reading, Grade 3	109
Table 54.	Distribution of Gain Scores Reading, Grade 4	109
Table 55.	Distribution of Gain Scores Reading, Grade 5	110

LIST OF TABLES (Cont'd)

		Page
Table 56.	Distribution of Gain Scores Math, Grade 3	110
Table 57.	Distribution of Gain Scores Math, Grade 4	111
Table 58.	Distribution of Gain Scores Math, Grade 5	111
Table 59.	Distribution of Gain Scores: Secondary Sample	112
Table 60.	Approximate Comparison of Selected ESAA Percentiles With Anchor Test Percentiles and Corresponding Publisher's Percentile Ranks on Total Reading Scores: Grade 4	113
Table 61.	Approximate Comparison of Selected ESAA Percentiles and Corresponding Publisher's Percentile Ranks on Total Reading Scores: Grade 5	114
Table 62.	Longitudinal and Cross-Sectional Comparisons of Reading and Math Achievement	120
Table 63.	Treatment-Control Comparisons of Academic Exposure	129
Table 65.	Differences in and Observed Gains for the Treatment Group Based on a Control Group Growth Curve	146
Table 66.	School and School Level Comparisons of Pretest-TC and Posttest-TC Correlations	147
Table 67.	ESAA Scale Score MANOVA Models	148
Table 68.	T-C Comparisons for the Longitudinal Sample Over the Entire Three Years of the Evaluation (ESAA Scale Scores)	152
Table 69.	Significant Regression Coefficients in Models Predicting Reading Achievement, Posttest Measures From Selected Predictors, and Pretest for Grades 3, 4, and 5	157
Table 70.	Significant Regression Coefficients in Model Predicting Mathematics Achievement Posttest Measures from Selected Predictors and Pretest for Grades 3, 4, and 5	158
Table 71.	Distribution of Scores Measuring the Extent to Which Instructional Programs Were Coordinated Throughout the School . . .	163
Table 72.	Distribution of Schools by Level of Program Coordination and School Success	165
Table 73.	Average Cost of the Principal's Time in Program Planning, Needs Assessment and Program Evaluation	169
Table 74.	Average Cost of the Principal's Time in Curriculum Development	169
Table 75.	Average Cost of the Principal's Time in Inservice Training .	169

LIST OF TABLES (Cont'd)

		Page
Table 76.	Average Overall Cost of the Time the Principal Spent in Instructional Management	170
Table 77.	Average Per-Teacher Cost of the Teachers' Time in Program Planning	170
Table 78.	Average Per-Pupil Cost of Teachers' Time in Reading and Math Instruction	171
Table 79.	Average Per-Pupil Cost for Equipment Associated With Reading and Math Instruction	172
Table 80.	Average Per-Pupil Cost of Instructional Aides Assisting Teachers With Primary Responsibility in Reading and Math Instruction	172
Table 81.	Average Cost Per Teacher for Inservice Training in Reading, Math, and in Teaching Disadvantaged Students	173
Table 82.	Average Hours Various Personnel Delivered Teacher Inservice Training by Type of School	174
Table 83.	Average Percentage of Inservice Delivery Time Attributed to On-Site Personnel (the Principal and Teachers)	175
Table 84.	Analysis-of-Variance of School Average Gain in Student Achievement	177
Table 85.	Intercorrelations Among Predictor and Outcome Variables in the 'Best' Model of School Average Gain in Student Achievement	177
Table 86.	Intercorrelations Among Resources Representing the Principal's Time in Instructional Management Activities	183
Table 87.	Average Per-Pupil Cost of Remedial Specialists Assisting Teachers With Primary Responsibility in Reading and Math Instruction	183
Table 88.	Average Per-Pupil Cost of Speech Therapists Assisting Teachers With Primary Responsibility in Reading and Math Instruction	183
Table 89.	Average Per-Pupil Cost of Media Specialists Assisting Teachers With Primary Responsibility in Reading and Math Instruction	184
Table 90.	Average Per-Pupil Cost of Counselors and Psychologists Assisting Teachers With Primary Responsibility in Reading and Math Instruction	184
Table 91.	Distribution of Schools Providing Counselors or Psychologists By School Success (With Mean Costs)	184

LIST OF TABLES (Cont'd)

		Page
Table 92.	Intercorrelations Among Predictor and Outcome Variables Used in Models of School Average Gain in Student Achievement . . .	185
Table 93.	Observer Sex and Ethnicity, and Evaluations of Teaching Practices	195
Table 94.	Comparisons of Original and Adjusted Items	197
Table 95.	Distribution of Teacher Satisfaction Scale Scores	221
Table 96.	Construction of the Teacher Satisfaction Measure	226
Table 97.	Comparison of Mean Teacher Satisfaction Scale Scores in Successful and Nonsuccessful Schools in Prior Year Achievement Gains (Outcome A)	227

LIST OF FIGURES

		Page
Figure 1.	Achievement Test Score Growth Patterns for Pilot Longitudinal Students: Reading	9
Figure 2.	Achievement Test Score Growth Patterns for Pilot Longitudinal Students: Mathematics	10
Figure 3.	Achievement Test Score Growth Patterns for Basic Elementary Longitudinal Sample Students: Reading	11
Figure 4.	Achievement Test Score Growth Patterns for Basic Elementary Longitudinal Sample Students: Mathematics	12
Figure 5.	Testing Design for Restandardization Study	16
Figure 6.	Hypothetical Distributions of Ability for Successive Age Groups	17
Figure 7.	Mean CTB Reading Scale Scores for Pre- and Posttestings in Grades 3 and 4	31
Figure 8.	Mean ESAA Reading Scale Scores for Pre- and Posttestings in Grades 3 and 4	32
Figure 9.	Profile of ESAA Scale Reading Scores	52
Figure 10.	Profile of ESAA Scale Math Scores	53
Figure 11.	Profile of CAT Scale Reading Scores	54
Figure 12.	Profile of CAT Scale Math Scores	55
Figure 13.	Longitudinal Profile of ESAA Scale Reading Scores for Each Decade (N=1574)	56
Figure 14.	Longitudinal Profile of SAA Scale Math Scores for Each Decade (N=1574)	57
Figure 15.	Longitudinal Profile of CAT Scale Scores for Each Decade (N=1574)	58
Figure 16.	Longitudinal Profile of CAT Scale Math Scores for Each Decade (N=1574)	59
Figure 17.	Cross-Sectional and Longitudinal Designs	116
Figure 18.	Longitudinal and Cross-Sectional Comparisons of Reading and Math Mean Achievement: Fall and Spring Testings	119
Figure 19.	Mean Achievement (By Year and Grade of Entry into Sample) of One-, Two-, and Three-Year Longitudinal Samples	121
Figure 20.	Matched, Unmatched, and Pseudo-Longitudinal Comparisons of Reading and Math Mean Achievement: Fall and Spring Testings	123
Figure 21.	Achievement Means for Longitudinal Students and Students Without Complete Data	125

LIST OF FIGURES (Cont'd)

		Page
Figure 22.	School Profiles	133
Figure 23.	Longitudinal Profile - Reading	150
Figure 24.	Longitudinal Profile - Math	151
Figure 25.	School Organizational Variables and Classroom Factors Related to Student Task Articulation	205
Figure 26.	Times at Which Achievement Gain and Teacher Morale are Assessed	211
Figure 27.	The Two Major Competing Hypotheses Explaining A Relationship Between Student Achievement Gains and Teacher Morale .	212
Figure 28.	The Two Major Competing Hypotheses Explaining A Relationship Between Student Achievement and Teacher Morale, Modified to Include Student Socio-Economic Status	217
Figure 29.	Zero-Order Correlations Among Variables in Hypothesis A . .	224

LIST OF EXHIBITS

		Page
Exhibit 1.	The Use of Figures 9-12 to Make Longitudinal Comparisons . .	70
Exhibit 2.	The Use of Figures 9-12 to Make Longitudinal Comparisons . .	71
Exhibit 3.	The Use of Figures 13-16 for Longitudinal Comparisons . . .	73

I. INTRODUCTION

Although the major activities of the national evaluation of ESAA Basic and Pilot Programs ended on November 30, 1977, System Development Corporation received a contract extension to perform several additional tasks related to the ESAA evaluation in the period from December 1, 1977 to March 31, 1978. One of those tasks involved the preparation of a non-technical overview of major findings and conclusions from the entire three-year study, to supplement the more detailed reports previously published. A second task was to prepare an overview, suitable for publication in a professional journal, of methodological issues and solutions in the ESAA evaluation. Those two tasks were completed, and the results have been documented elsewhere.*

The present report focuses on a set of supplemental analyses that were performed on portions of the ESAA evaluation data during the extension period. The goals of these analyses were to explore additional relationships in the data that might help to inform program policy; to confirm and/or further explicate some of the findings reported earlier; and to put some of the study results, particularly those concerned with student acheivement, in a form that might be useful to individuals directly concerned with the operation or evaluation of educational intervention programs such as ESAA.

The remainder of this report is organized into three sections. First, the goals and design of the ESAA evaluation are briefly reviewed. The second section describes the results of additional analyses performed on data from the overall ESAA evaluation, i.e., from the sample used for the program impact study. Finally, results of supplemental analyses of data from a smaller in-depth study sample are presented.

II. OVERVIEW OF ESAA AND OF THE EVALUATION

The ESAA Basic Program is a desegregation assistance program in which grants are awarded to eligible school districts to encourage the reduction of minority-

*A comprehensive list of reports describing the goals, methodology, and results of the three years of evaluation is presented at the end of this report.

group isolation, to meet the special needs incident to the elimination of segregation and discrimination, and to assist elementary and secondary school children in overcoming associated educational disadvantages. In contrast to Basic grants, Pilot Project grants are awarded to local education agencies for unusually promising projects designed to overcome the adverse effects of minority-group isolation by improving the academic achievement of children in minority-isolated schools, i.e., schools with 50 percent or greater minority enrollment. Thus, the Pilot Program is essentially a compensatory education program.

The ESAA evaluation project had two major components: a large-scale impact study involving national samples of schools with ESAA Basic or Pilot projects, and a more intensive, in-depth study involving a subset of those schools, designed to identify specific program approaches differentiating successful from unsuccessful projects. For convenience, those two components are referred to in this report as the 'overall evaluation' and the 'in-depth study'.

The design of the 'overall evaluation' involved the selection of two demographically similar, ESAA-eligible schools from each sample district. One of the two schools was randomly selected for the treatment group, i.e., to receive ESAA funds, while the other, denied ESAA funding, became the control school. Students were randomly selected across classrooms at grades 3 through 5 in elementary schools, and grades 10 through 12 in secondary schools. For each of three school years, 1973-76, comparable student, school, program, and outcome data were collected in both treatment and control schools. Achievement tests were administered each year as early as possible after the beginning of classes--in many cases this was not until about two months after school opening--and again toward the end of the school term. In addition, questionnaires were administered each year to superintendents, ESAA coordinators, district business managers, principals, teachers, and students. Analyses for the overall study focused on ESAA program impact (effect on student achievement as determined by comparisons of gains in treatment and control schools), and on relationships between various school and program characteristics and student achievement gains. Table 1 shows changes in the number of paired treatment and control schools during the three evaluation years, as well as the number of schools that were in the study over the full three years, i.e., the longitudinal sample.

Table 1. Sample Changes Over Years

Sample	Schools in Treatment-Control Parts			
	Year One	Year Two	Year Three	Longitudinal
Basic Elementary	74	68	44	34
Basic Secondary	24	40	28	16
Pilot Elementary	36	40	30	16

At the student level, the total sample sizes for the yearly impact analyses were 19,573 in the first year, 17,297 in the second year, and 9,541 in the third year. The longitudinal sample, consisting only of students who were in the first year sample, at grade 3 or grade 10, and who were still in the study for third-year posttesting at grade 5 or 12, included a total of 1,383 students.

The 'in-depth study' was conducted during the second and third years of the overall evaluation (school years 1974-75 and 1975-76). In both years, the overall goal was to identify the program and contextual characteristics that differentiated most clearly and consistently between successful and unsuccessful elementary schools. For this purpose, intensive classroom observations and staff interviews were conducted in subsamples of elementary schools (24 schools in 1974-75 and 26 schools in 1975-76) selected from the overall evaluation sample on the basis of their wide variations in program effectiveness, i.e., in the schools' ability to increase students' reading and math achievement.

III. ADDITIONAL ANALYSES OF DATA FROM OVERALL EVALUATION

The analyses and related activities discussed in this section were based on data from the overall evaluation, as defined above. The first set of activities described below was concerned exclusively with the standardized achievement test data. Those activities included some new analyses, but consisted primarily of efforts to re-format and document the test results in a manner that would make them more informative and useful to individuals operating or evaluating ESAA-like programs. The second part of this section describes efforts to derive new

indices and composite variables from the existing data, in anticipation that these new indices might reveal important additional relationships in the data, or at least clarify or strengthen the relationships previously reported.

A. ACTIVITIES TO INCREASE INTERPRETABILITY AND UTILITY OF ESAA TEST DATA

Three subtasks were performed to provide clearer and more useful information about the test performance of sample students in ESAA districts. First, the restandardization of the test data to provide greater interpretability of the students' growth patterns was documented more fully. (Although the basic restandardization procedures had been discussed in an earlier report, the material presented below describes in greater depth the problems of test score distributions that led to the restandardization effort, and the improvements in the growth curves that resulted from that effort.) Second, summary tables and graphs showing the test scores for the ESAA sample students are presented in a form designed to facilitate their use as baseline or comparison data for assessing the achievement of students in local ESAA projects, or in other programs having students with similar demographic characteristics. Finally, the achievement growth curves developed from true longitudinal data are compared with the curves developed from a quasi-longitudinal design (using data cutting across grade levels but collected at a single point in time); it was felt that such a discussion might help to clarify issues related to the comparability of growth data generated by those two different techniques.

1. RESTANDARDIZATION OF ESAA ACHIEVEMENT TEST DATA

A restandardization of portions of the 1970 edition of the California Achievement Test (CAT) was conducted as part of the three-year longitudinal evaluation of the ESAA Basic and Pilot programs. This restandardization was an attempt to remedy two methodological problems in testing disadvantaged, low-achieving students, that became apparent during the first two years of the study. The remainder of this volume focuses on the methodological problems, the procedures used to rectify them, and the results obtained from the application of these procedures. Preceding these discussions, a brief description of the ESAA evaluation design for administration of the California Achievement Test (CAT) is provided as background information.

a. ESAA Test Administration Design

The design of the ESAA evaluation involved the administration of achievement tests to sample students (grades 3, 4, and 5 at the elementary level, and grades 10, 11, and 12 at the secondary level) in treatment and control schools near the beginning and the end of each school year. At the elementary school level, the 1970 edition of the California Achievement Test (CAT) was used as the academic achievement outcome measure. Measures of performance for both reading (Vocabulary and Reading Comprehension) and Mathematics (Computations and Concepts) were included in this test.* These tests were available in a series of sequenced overlapping test-levels spanning the grade range of K-12. The published materials indicated that no single test-level was appropriate for students in grades 3, 4, and 5. Rather, CAT level-2 was recommended for use in grades 2.0 through 4.9 and level-3 for grades 4.0 through 6.9. It was necessary, given these recommendations, that more than one test-level be used with the elementary students. This meant that three-year longitudinal students would have to change test-level at one point during the course of the study. It was decided that this point would be at the end of grade 3. Thus, students would be administered the level-2 test at both testing points in the 3rd grade and, starting with the fall testing in the 4th grade, would receive the level-3 test for the remaining testing points. This change in test-level was not viewed as a serious problem as the publisher had provided vertical scales spanning all the test-levels in the battery which allowed direct comparison of performance regardless of test-level. More discussion of the nature of these scales will be presented later.

At both test-levels, two alternative forms (Form A and B) of the test were available. These were administered to the sample students in a counter-balanced design with random halves of the sample students within each participating school and grade receiving one form on the first occasion and the remaining form on the

*A special version of the mathematics computations subtest was used in the ESAA evaluation. This version omitted problems involving fractions. The test publisher provided national norm data for this special test version. The make-up of this test is discussed at length in a document by Ozenne, et al. 1975.

second occasion. The alternate forms had been equated by the test publisher so that identical percentile standings on the two test forms would result in identical vertical scale scores.

The testing design was considered to be an optimum strategy for measuring the academic growth of sample students over the six testing points in the evaluation. The use of alternate forms precluded the practice effects from one testing to the next and, while the change in test-level was not viewed as an asset, it was not considered a liability because of the availability of the vertical scales. However, the results of the first two years of the evaluation indicated that non-trivial problems might exist both in the form-to-form equating and, more significantly, in the vertical scaling of the CAT for our sample.

b. Problems in the Vertical Scaling of Publisher's Norms

Vertical scales are often available with major published standardized achievement tests. For any specified test battery, a vertical scale constitutes a single continuum or scale that spans all the test-levels and forms available as part of that test battery. The creation of such a scale requires successive scaling of performance for different test-levels. This is usually obtained through the use of overlap grade-or-age levels where two or more levels of the test are administered. The performance of the common group on the different test-levels provides the necessary linkage for creation of the vertical scale. (The entire scaling process is described in detail in a later section of this volume.) The advantage of such a scale is that "The percentile rank corresponding to a given scale score does not depend on the form or level (of the test) taken", (Tiegs and Clark, 1971). Further, such a scale provides the vehicle for the study of long-term growth over periods which require the use of more than one test-level.

In the ESAA evaluation, the vertical scale (called Achievement Development Scale Scores or ADSS by the test publisher) was used as the primary raw data value for all analyses involving comparisons, aggregations, or academic change. The ADSS score became particularly important in the second and third years of the study,

when change or growth over time was being measured and contrasted; and it was during the analysis of the second-year longitudinal data (Coulson, et al., 1976) that the first evidence appeared of a possible problem in the vertical scaling. The impact analysis of the second-year longitudinal data examined the growth pattern across two years of data collection (four testing points). Two different cohorts were used in the investigation. The first cohort consisted of students who were in the 3rd grade during the first year of the study and were in the 4th grade in the second year. This group is referred to as the grade 3/4 cohort. The other cohort, the grade 4/5 cohort, included students who started the first year of the study in grade 4 and were in grade 5 during the second year. The results of the impact analysis for the two groups are presented in Table 2, and shown graphically in Figures 1-4.

It can be seen that the growth curve for the grade 3/4 cohort for the period between the end of grade 3 and the beginning of grade 4 was markedly different from the positively accelerating slope found to exist between the remaining pairs of points for this cohort. The growth between these two points was negative (Pilot sample) or nearly zero (Basic Elementary sample). This result would not be expected from either educational theory or vertical scaling, both of which assume positive growth over time.

The possibility that this result could be due entirely to normal summer inactivity was discounted for two reasons. First, the period between the two testing points was approximately six months in duration and included about three months of instructional activity. Thus, some reduction in extent of the positive growth slope might be reasonable, but not a zero or negative slope. Second, the grade 4/5 cohort, which data was collected on the same time schedule as the 3/4 cohort, did not show results consistent with those for the grade 3/4 cohort. No evidence of the anomaly was apparent in the growth curve displayed by the grade 4/5 cohort during the period between 4th-grade posttesting and 5th-grade pretesting. Rather, this curve was, as would be expected of a group learning over time, positively increasing. (It is interesting to note also that, with the exception of the anomalous part of the 3/4 curve, the growth patters for the two cohorts were very similar.)

Table 2. Mean Achievement Scores for Students Who Were in the ESAA Evaluation Sample for Two Years

Program	Cohort Grade Year 1/ Year 2	Treatment Control	Year 1				Year 2			
			Pretest		Posttest		Pretest		Posttest	
			Reading	Math	Reading	Math	Reading	Math	Reading	Math
Pilot	3/4	Treatment	309.38	280.76	330.67	313.68	327.05	314.78	349.19	343.14
		Control	315.55	287.59	337.58	318.83	337.77	323.46	360.75	350.24
	4/5	Treatment	323.40	310.64	347.09	338.61	367.73	351.26	377.66	371.44
		Control	322.56	310.62	350.46	342.68	369.23	359.96	387.27	383.93
Basic	3/4	Treatment	332.57	296.67	356.13	330.78	357.69	334.61	380.14	363.05
		Control	331.69	295.36	353.52	331.02	358.96	335.32	377.44	358.91
	4/5	Treatment	356.22	332.65	381.40	363.20	399.99	380.03	415.05	399.18
		Control	352.80	331.60	379.00	359.60	395.67	375.82	409.58	393.15

Figure 1. Achievement Test Score Growth Patterns for Pilot Longitudinal Students: Reading

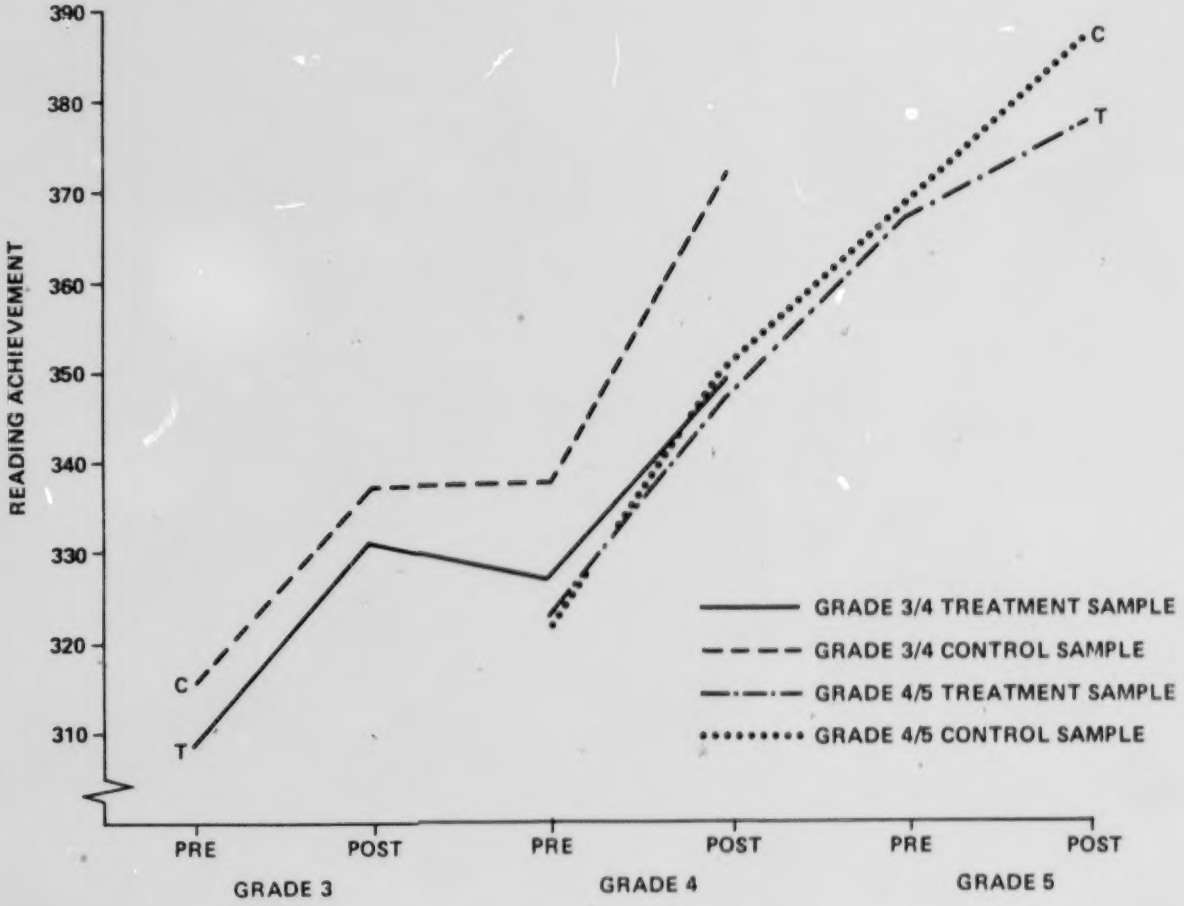


Figure 2. Achievement Test Score Growth Patterns for Pilot Longitudinal Students: Mathematics

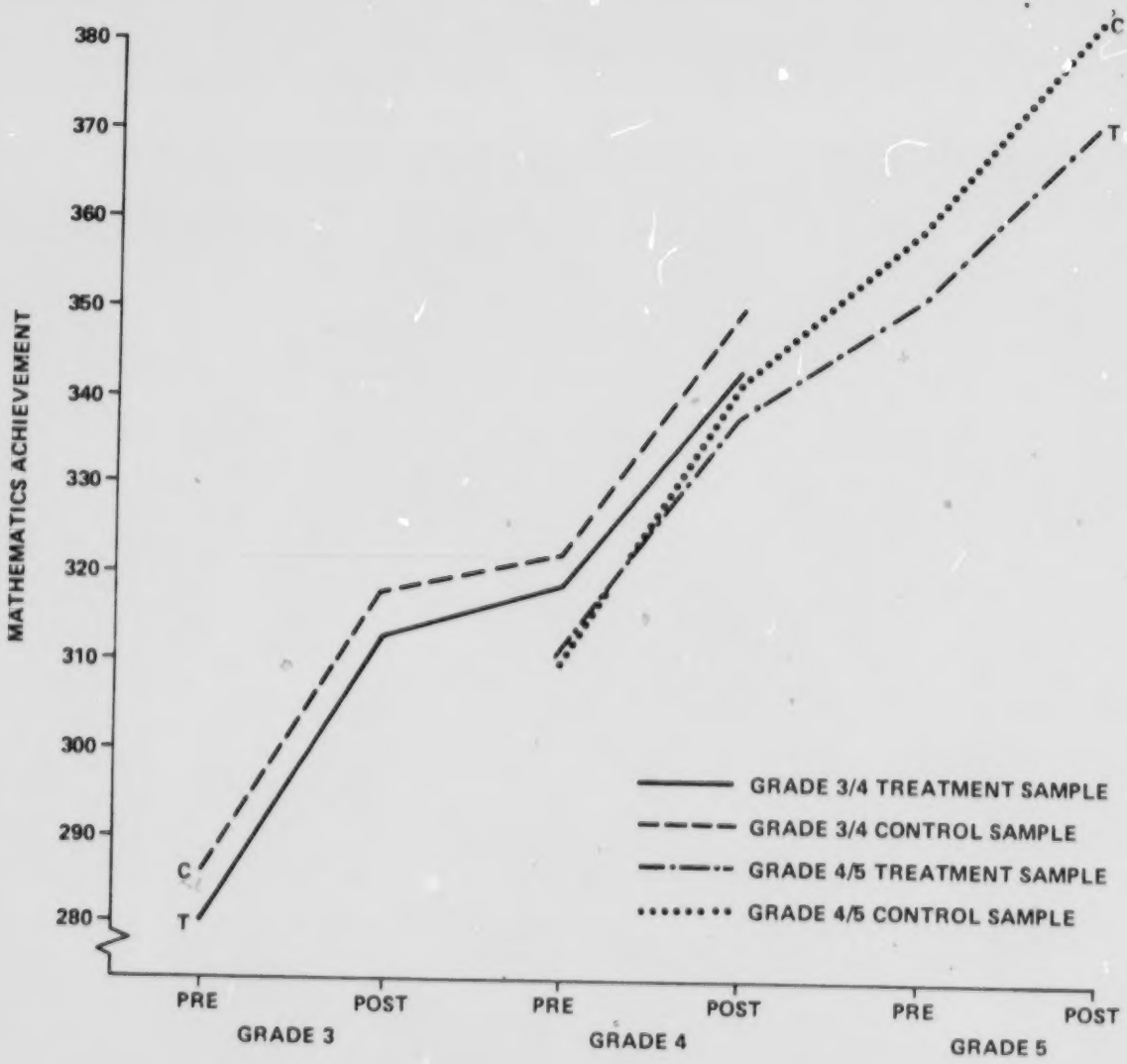


Figure 3. Achievement Test Score Growth Patterns for Basic Elementary Longitudinal Sample Students: Reading

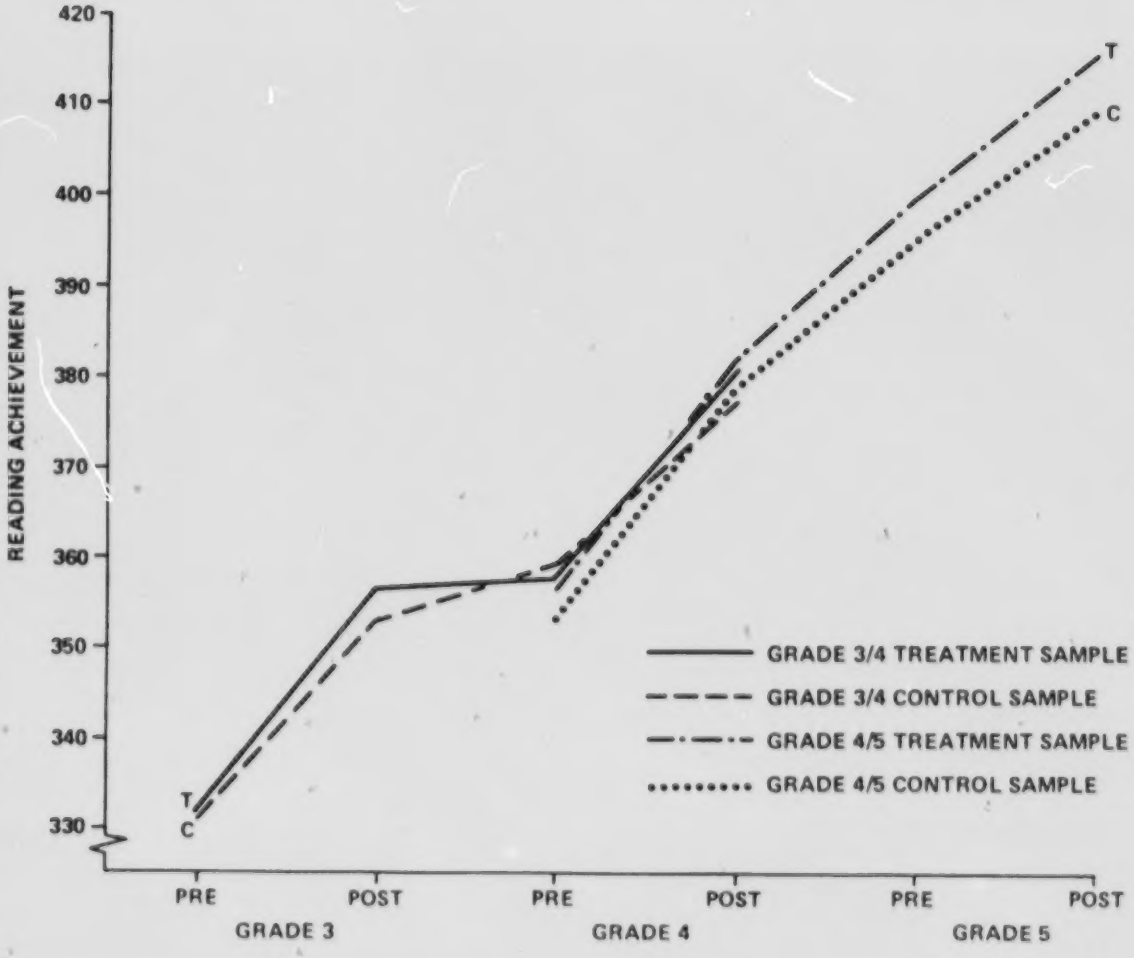
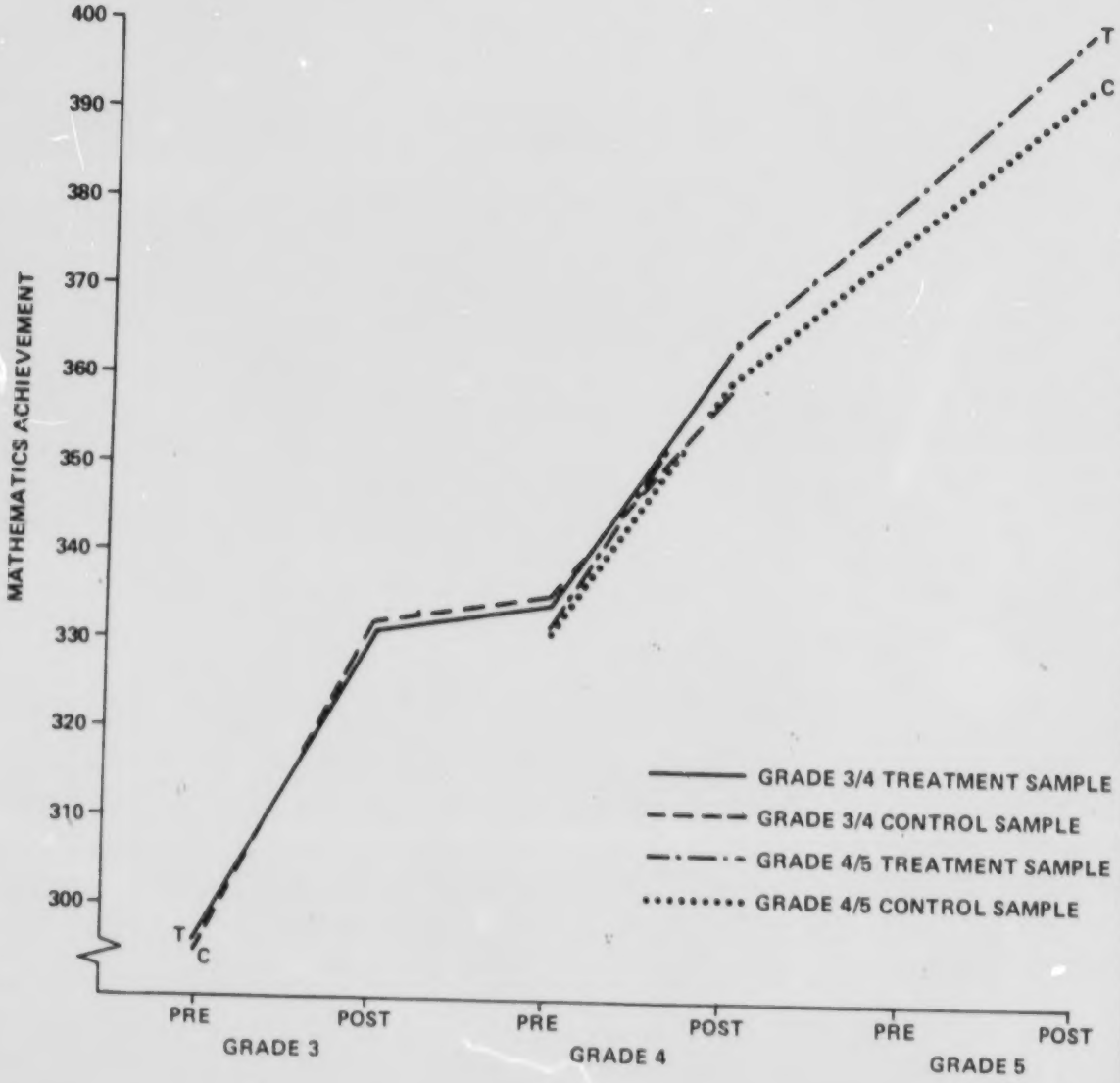


Figure 4. Achievement Test Score Growth Patterns for Basic Elementary Longitudinal Sample Students: Mathematics



At this point, the data were further examined to determine whether the apparent anomaly in the grade 3/4 cohort resulted from artifacts or from a true difference in growth patterns between the 3/4 and 4/5 cohorts. Descriptive evidence did not suggest any substantive differences between the grade 3/4 cohort and the grade 4/5 cohort. Also, since the phenomenon in question occurred in both the Pilot and Basic samples, it was not unique to a particular sample and could hardly be considered a fluke. Indeed, the only unique aspect of the phenomenon was that it occurred at the one point in the evaluation design where the students in the grade 3/4 cohort changed test-levels. As mentioned previously, the 3rd-grade results for that cohort were based upon performance on the level-2 test, while the 4th-grade fall testing employed the level-3 test. The grade 4/5 cohort results, by comparison, were based entirely on level-3 results. The fact that the performance of the grade 3/4 cohort had a positively accelerating slope between testing points when the same test-level was used also seemed to indicate that the phenomenon might be restricted to some problem in the articulation of the two different test-levels.

These facts led to an examination of the information provided by the publishers about the vertical scales. The initial inspection of the published information for the normative population indicated that for the nine decile points (the decile point defines the performance of successive groups of 10 percent of the group) the same scale score would be assigned regardless of whether a student had taken the level-2 or level-3 test. Also, if one assumed that students in the normative population would maintain percentile rank from one testing point to the next, the resultant scale scores should increase in value as would be expected by the assumption of positive cognitive growth. The only conclusion to be drawn from this investigation was that some interaction existed between the test-level articulation and the nature of the ESAA sample, such that the vertical scaling provided by the test publisher was not functioning as expected for the ESAA evaluation data.

The presence of such an interactive effect with the ESAA students seemed not unreasonable, considering the special characteristics of the ESAA population. The ESAA students as a group, and particularly in the Pilot Program, tended to

be educationally disadvantaged, primarily from minority heritage and academically low-achieving. They were certainly dissimilar in many important demographic characteristics from the students constituting the great bulk of the normative population used by the publisher to establish the vertical scales. It seemed plausible, then, that differences in the growth and performance patterns of the ESAA students and the publisher's normative group might have resulted in scaling and equating anomalies, since such scaling and equating operations are based upon the distributional and performance patterns of the normative group. To remedy this condition, it was decided that a rescaling of the CAT, specific to the ESAA students, should be conducted and used in future ESAA evaluation analyses.

c. Problems in Equating Alternate Forms of Publisher's Norms

There were also difficulties in the equating of the two alternate forms, though these were not as apparent as the problems with the vertical scaling. As mentioned previously, the publisher had provided conversions based upon the normative population so that an individual would receive the same vertical scale score regardless of test form. In other words, identical percentile rank performance on the two test forms would produce the same scale score. However, the results of the first two years of the evaluation seemed to indicate that this might not be true for our sample.

Two different pieces of evidence lead to this conclusion. First, the raw ADSS gain (posttest minus pretest) scores were larger, on average, for those people who took Form B in the fall and Form A in the spring, than for people who had been administered the tests in reverse order. Since the students had been randomly assigned to test form, this condition was not consistent with the expectation of random sampling. Secondly, the corresponding decile performance of the ESAA sample (all three grades were included in this comparison) was compared across the two test forms. The results indicated that the vertical scale scores were not the same for the two groups (test forms), and that the direction and magnitude of the differences varied with test-level and decile level. That is, at some deciles the Form A produced higher ADSS scores, while at other deciles Form B produced higher scores. This evidence suggested that the alternate-form

equation might not be accomplishing its proposed function, and the resulting equation error might be leading to an overestimate or an underestimate of the basic data values of the ESAA evaluation (raw ADSS gain scores), depending upon the test form sequence and the decile level of the student. Clearly, the introduction of such variation could hamper the attempt to isolate the program effects of ESAA.

For the above reasons, a restandardization study to rescale and renormalize the achievement measures was designed to meet the following objectives:

- To construct a common metric scale, so that a unique scale score would be derived for students of equal achievement regardless of test-level.
- To equate the two alternate forms so that scale score estimates of student performance would not be affected by the test form utilized to measure the ability.
- To provide percentile rank norms for a population with the characteristics of the ESAA participants.

A random group design was implemented with the data from third-year sample students. The random group design (Angoff, 1971) can be utilized when it is necessary to create a single vertical scale for a series of sequenced 'overlapping' test-levels. Overlapping test-levels are tests that have recommended grade ranges which comprehend the same grade. As mentioned earlier, level-2 of the CAT was recommended for grades 2.0 through 4.9, and level-3 was appropriate for grades 4.0 to 6.9. Grade 4 was the 'overlap' grade for these two test-levels. The random group design requires that the sample students at this 'overlap' grade be randomly assigned to one or the other test-level. This meant that the 4th-grade sample students were randomly and evenly divided into groups receiving the two test-levels. Figure 5 presents graphically the final testing design for the restandardization study.

Figure 5. Testing Design for Restandardization Study

	Level-2		Level-3	
	Form A	Form B	Form A	Form B
Grade 3	X	X		
Grade 4	X	X	X	X
Grade 5			X	X

As indicated in Figure 5, each sample of students at a grade level was further divided in half, with each half being randomly assigned to one of the two alternate test forms. The design resulted in two testing groups at each of grades 3 and 5, and four testing groups at the 4th-grade level. The assignment of pupils to test form, and in the case of grade 4, to test-level, was random within schools and independent, school to school.

d. Analytic Model Used to Restandardize

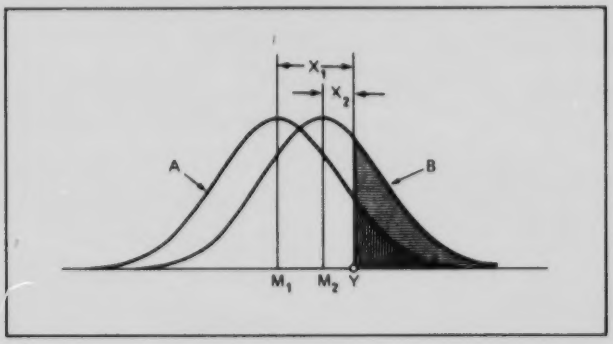
In the present case, two separate methods were used to accomplish the study's objectives. The construction of the common metric was performed using Thurstone's (1925) Absolute Scaling Method. The equation of the alternate test forms was accomplished by the equipercentile method, (Angoff, 1971). The assumptions and basic methodology of each of these procedures are described below.

Thurstone (1925) presented the method of Absolute Scaling as a technique for scaling items when these have been applied to samples from successive age or grade groups. As Thurstone said, "My present method applies to test items that can be graded right or wrong, and for which separate norms are to be constructed for successive age-or-grade groups." He indicated that the method was applicable to tests that give a numerical score. Examples of this procedure's application can be found in Thurstone (1933) and Teigs and Clark (1970).

Regardless of the application, the Thurstone method assumes a single undimensional scale on which the various age (or grade) groups are normally distributed. It

also assumes that these distributions are displaced from each other in terms of their means, but it does not assume that they have equal dispersions. The situation assumed by the Thurstone technique can be thought of as a series of frequency distributions, all drawn on the same base line, that truly represent the distributions of ability or achievement of children of successive chronological age. Each test item (or score) can be located in this base line. The curves must be drawn so that the proportion of cases to the right of a test question will correspond to the actual number of cases answering the item correctly. This is depicted in Figure 6.

Figure 6. Hypothetical Distributions of Ability for Successive Age Groups



In this figure M_1 and M_2 represent the mean values of the two populations, and Y represents the location of a particular test score. The shaded portions of the distributions indicate the percentages of cases achieving at least that score or a higher score. The distances of point Y from the means, shown by X_1 and X_2, can be expressed in terms of standard scores for the two distributions. Since the two standard scores correspond to the same data point, they can be thought of as being equal. Thus, if the data fit the Thurstonian model, it is possible to translate the successive groups onto a single scale. This is accomplished through the necessary linear relationship between the normal deviates of the successive groups. Specifically, it can be seen that the point Y represents $M_1 + X_1\sigma_1$ for group 1, and $M_2 + X_2\sigma_2$ for group 2. Since this is the same point, it may be expressed as:

$$M_1 + X_1\sigma_1 = M_2 + X_2\sigma_2 \qquad (1)$$

or

$$X_1 = X_2 \frac{\sigma_2}{\sigma_1} + \frac{(M_2 - M_1)}{\sigma_1} \qquad (2)$$

where M_1 and M_2 are the mean values of the two groups in terms of the scale, X_1 and X_2 are the normal deviates, and σ_1 and σ_2 are the dispersions (standard deviations) of the groups in terms of the scale values. The calculation of the linear constants in terms of observed values is:

$$T_2 = \frac{S_1}{S_2}\sigma_1 \qquad (3)$$

or

$$M_2 = \sigma_1 (M_1 - \frac{S_1}{S_2}) + M_1 \qquad (4)$$

where S_1 and S_2 are the dispersions of the observed normal deviates, and M_1 and M_2 are the means of these deviates. It is apparent from equations (3) and (4) that the mean and standard deviation of one group must be selected as an anchor point so that all other groups can be translated onto the scale with reference to the distribution of this group. The choice is arbitrary and usually is decided for convenience. Following selection of an anchor group, the mean and standard deviation of the remaining groups are determined by (3) and (4). The scale score for each test score within a group is computed by using:

$$Y_1 = M_1 + X_1\sigma_1 \qquad (5)$$

This procedure results in a single scale in which each item or score has a unique value within a group. If more than one group has taken the same test, the group scale values are averaged to arrive at final score values for each raw score.

While the Thurstone procedure is appropriate for the construction of a common metric, it cannot fully satisfy the current study's need for equating alternate test forms. The use of the Thurstone procedure hopefully ensures (1) that the

relationship between the normal deviates of successive groups is linear and (2) that the underlying ability distributions for each group are normal. These assumptions are valid only with common sets of items. Because of this requirement, the link between the level-2 and level-3 scales must be provided by a different technique. Equipercentile equating was chosen as the vehicle for this linkage.

Angoff (1971) describes equipercentile equating in detail. Essentially, the method produces a nonlinear transformation of one set of scores onto a second set, and is based upon the determination of scores on two tests that have the same mid-percentile rank. The result of this process is that a person with a specific ranking in the population (percentile rank) will receive the same score value regardless of which test form he actually took. This procedure does not require common sets of items. Therefore, it can be used both for establishing the linkage between level-2 and level-3 scale scores and for the equating of the alternate test forms within each level.

These two methods (Thurstone's Absolute Scaling technique and equipercentile equating) provided the mechanism for the accomplishment of the restandardization objectives. It was believed that these techniques would produce the most reliable and technically sound results, while observing all practical limitations. The implementation of this methodology is described below, and results are presented in a subsequent section.

e. Restandardization Procedures

This section describes the procedures used to apply the methodology described above to data collected at the beginning of the third year of the ESAA evaluation. Three general comments should be made before covering the actual steps in detail. First, each of these steps was performed separately for the six scores derived from the CAT used in the ESAA evaluation: vocabulary, reading comprehension, total reading, mathematics concepts, mathematics computations, and total mathematics. Second, parallel but separate scaling was conducted for the Form A and Form B tests. Third, the Form A and Form B were equated using the new scale scores rather than the raw scores. (The reason for the last two steps will be discussed in the results section.)

After the raw data were validated, raw score frequency distributions were compiled. Separate distributions were formed by grade, test-level, test form, and subject matter. The mid-percentile ranks were computed corresponding to each raw score point within these distributions. The distributions were then smoothed by means of an analytic technique employing rolling weighted averages, developed by Cureton and Tukey (1951).

Each smoothed distribution was normalized by generating the normal deviate corresponding to each of the smoothed mid-percentile ranks as if it were from a normal distribution. In the few cases where the mid-percentile ranks were zero or unity, linear extrapolation was used to generate new mid-percentile ranks that would produce reasonably normal deviates. These normal deviates supplied the necessary information to begin the absolute scaling procedure. First, the normal deviates from adjacent grade levels were plotted against each other. This resulted in two plots: one for grade 3 level-2 versus grade 4 level-2, and another for grade 5 level-3 versus grade 4 level-3. The linearity of these plots was ascertained both visually and through correlation coefficients. The correlations were all over .99.

Using the computational formulas (3) and (4) from the previous section, the relationship between the means and standard deviations of the adjacent grades was ascertained. The grade 5 level-3 group was chosen as the anchor group for the scaling. A mean value of 600 and a standard deviation of 100 were assigned so that the scale scores would range from approximately 0 to 999.

While the scaling procedure provided the mean and standard deviation within test-level, equipercentile equating was used to link the level-3 and level-2 scales. This was accomplished by assigning the scale mean and standard deviation of the grade 4 level-3 group to the grade 4 level-2 group. This was, in effect, an equipercentile equating, since both distributions had been normalized.

After each grade's mean and standard deviation on the new scale was determined, formula (5) was used to generate a scale value for each test score. This process was separately performed at each grade level and resulted in two scale

estimates for each test score. The final scale values were derived by averaging these two grade-level estimates.

The equating of Form A to Form B was initiated by generating scale score distributions by test-level. That is, separate frequency distributions of scale scores were generated by combining the two grade-level distributions at each test-level. Mid-percentile ranks for each test-level were computed and smoothed as described above. This was repeated for each test form. The Form A scale score was then determined. In certain cases, these values were linearly extrapolated to cover all the initial raw score data points. The final scale scores for use in the ESAA evaluation can, thus, be considered comparable across test-level or test form.

f. Results of Restandardization

This section presents the results obtained from the analysis of the restandardization data. Discussion of these results has been organized into three parts: description of raw score performance, results of the scaling, and results of equating the forms.

(1) Description of Raw Score Performance

Table 3 lists the numbers of items in the four CAT subtests and in the reading and mathematics total score.

The number of items is the same across the two alternate test forms within a given test-level. In all cases, scores were simply the number of items answered correctly by the student, as no correction for guessing was applied.

Table 3. Number of Items in CAT Subtests

	Level 2		Level 3	
	Form A	Form B	Form A	Form B
Vocabulary	40	40	40	40
Comprehension	45	45	42	42
Total Reading	85	85	82	82
Mathematics Concepts	30	30	25	25
Mathematics Computations	72	72	48	48
Total Mathematics	102	102	73	73

Test results for the third-year students are presented in Table 4. Separate summaries have been provided for each of the eight testing groups. The mean and standard deviation of raw score performance have been provided for all six raw scores. Reference to Table 3 allows a comparison of the mean performance against the maximum possible scores.

It should first be noted that, as expected, the level of performance was uniformly below the mean of the national normative group. In general, the ESAA students' performance was at about the 30th percentile relative to that group. Second, there was a noticeable difference in raw score performance level between Form A and Form B. This confirmed the suspicion that Form B of the CAT was a more difficult test. Third, differences in the shapes of the 4th-grade score distributions at the two test-levels indicated a large discrepancy in test difficulty across test-level as well as test form. This finding may be important in our understanding of the existence of level articulation problems in the original CAT scale scores.

(2) Results of the Scaling

Mid-percentile ranks were computed for each raw point in the grade-level score distributions. These were then smoothed, using the Cureton and Tukey (1951) analytic procedure, to provide grade-level percentile norm tables for the different test-levels and forms. The results can be found in another section of this report (Section III,A,2).

Table 4. Raw Score Performance of ESAA Third-Year Students

Grade	Level/ Form		Reading			Mathematics		
			Vocabulary	Comprehension	Total	Concepts	Computation	Total
3	Level 2 Form A	$\bar{X}$	28.08	22.87	50.95	17.94	42.90	60.84
		S.D.	8.78	10.90	18.57	6.11	13.32	17.82
		N	1475	1475	1475	1475	1475	1475
	Level 2 Form B	$\bar{X}$	21.79	23.54	45.34	16.60	35.12	51.72
		S.D.	8.66	10.48	17.21	5.91	11.07	15.39
		N	1486	1486	1486	1486	1486	1486
4	Level 2 Form A	$\bar{X}$	31.69	28.54	60.23	21.84	56.83	78.68
		S.D.	7.94	10.96	17.90	6.50	13.59	18.39
		N	797	797	797	797	797	797
	Level 2 Form B	$\bar{X}$	25.27	28.63	53.90	20.51	38.03	68.54
		S.D.	8.98	10.40	17.83	5.92	14.82	13.52
		N	794	794	794	794	794	794
	Level 3 Form A	$\bar{X}$	13.84	14.89	28.72	9.21	18.76	27.96
		S.D.	7.56	6.10	13.28	9.87	8.37	12.25
		N	809	809	809	809	809	809
	Level 3 Form B	$\bar{X}$	11.57	13.01	24.58	8.45	15.68	24.13
		S.D.	7.24	6.10	12.27	4.75	6.45	10.11
		N	804	804	804	804	804	804
5	Level 3 Form A	$\bar{X}$	18.18	18.26	26.44	12.19	25.11	37.30
		S.D.	8.20	7.54	15.45	5.08	9.92	14.07
		N	1737	1737	1737	1737	1737	1737
	Level 3 Form B	$\bar{X}$	15.47	15.78	31.25	11.34	20.79	32.14
		S.D.	8.52	6.95	14.49	5.60	7.57	12.11
		N	1727	1727	1727	1727	1727	1727

As mentioned previously, the 5th-grade level-3 test performance was used as the anchor for the Absolute Scaling, and the mean and standard deviation of this group were designated as having scale values of 600 and 100, respectively. The remaining testing groups' means and standard deviations were translated onto this scale. The results of this operation, presented in Table 5, indicate two things. First, the average 3rd-grade test performance is approximately one scale standard deviation below that of the anchor group for reading. This difference widens to one and one-half standard deviation for mathematics computation and mathematics total. Second, the independent scaling of the two test forms produced fairly similar results, with the possible exception of the 3rd grade. Third-grade performance on Form B was consistently closer to that of the anchor group than performance on Form A. This probably resulted from the increased difficulty of Form B, resulting in less variability in performance.

The scale value for each raw score point was computed, using the tabled scale means and standard deviations. This procedure yielded a grade-level scale score value for each raw score point. The final scale-score-to-raw-score conversion was obtained by averaging the two grade-level estimates within each test-level. The grade-level estimates and final score values are contained in another report (Doherty, 1977).

It is important at this time to consider whether the restandardization activities remedied the two problems to which they were directed. As indicated earlier, one motive for restandardization was to eliminate any effects of test-level articulation. The longitudinal characteristics of both the publisher's and the new ESAA scale scores were examined to determine whether the new scale scores reduced articulation problems. Pupil progress was 'tracked' at constant percentile ranks (the decile points between 10-90 were selected as being representative) across the six testing points that characterized the three-year ESAA evaluation design. This was done both for the national norm group, i.e., decile rank defined in terms of the national norming sample, and for the ESAA evaluation sample used in the restandardization study. This produced two sets of scores across the six testing points.

Table 5. Scale Means and Standard Deviations

		Reading			Mathematics		
		Vocabulary	Comprehension	Total Reading	Concepts	Comprehension	Total Mathematics
Fifth Grade Level 3	Form A	600 (100)	600 (100)	600 (100)	600 (100)	600 (100)	600 (100)
	Form B	600 (100)	600 (100)	600 (100)	600 (100)	600 (100)	600 (100)
Fourth Grade Level 3	Form A	549.04 (97.64)	550.65 (97.51)	544.28 (98.48)	540.28 (98.77)	533.95 (92.45)	530.84 (91.88)
	Form B	546.39 (99.02)	555.20 (106.06)	549.68 (100.70)	545.39 (93.83)	533.99 (94.03)	533.95 (94.02)
Fourth Grade Level 2	Form A	549.04 (97.64)	550.65 (97.51)	544.20 (98.48)	540.28 (98.77)	533.95 (92.45)	530.84 (91.88)
	Form B	546.39 (99.02)	555.20 (106.06)	549.68 (100.70)	545.39 (93.83)	533.99 (94.03)	533.95 (94.02)
Third Grade Level 2	Form A	505.81 (90.72)	498.44 (97.74)	489.31 (106.39)	482.54 (87.84)	447.72 (73.98)	449.11 (75.43)
	Form B	509.72 (88.45)	510.97 (91.54)	505.66 (86.82)	496.35 (79.46)	457.11 (76.64)	461.49 (76.77)

Tables 6 and 7 present the results of reading performance for both samples. At each testing point, the raw score, the CTB scale score, and the ESAA scale score are reported by decile rank. Examination of the decile data for the national normative sample, shown in Table 6, indicates that the change in scores is positive between all pairs of testing points. This pattern is maintained regardless of score type* (raw, CTB, or ESAA code) and seems congruent with the notion that achievement should increase with time. It is interesting to note that the magnitude of change in CTB scale scores is not constant across the deciles. Scores at or below the 30th percentile consistently change less than those at higher deciles across the six testing points. This phenomenon is consistent with the gradual increase in variance in such distributions over time. It is particularly striking for the period between the end of grade 3 and the beginning of grade 4. For this period, the changes in CTB scale scores for the 10th, 20th, and 30th percentiles are 6, 6.5, and 9.85, respectively, compared with an average change of 14.1 at the other six deciles. This would seem to indicate that the CTB norms are not reflecting a linear growth pattern and that scores below the 30th percentile--the score range in which most ESAA sample students placed, are growing less than the remainder of the distribution.

A different pattern is shown in Table 7, which contains scores based on percentile ranks computed within the ESAA evaluation sample. The table's interpretation is that the 10th percentile, for example, is the score value at or below which 10 percent of the ESAA third-year students performed. Three things are evident in these results. First, the performance of the ESAA-defined decile group is uniformly lower than that of the corresponding decile groups from the national norms. Second, the average amount of change between ESAA students' scores at successive testing points is less than that for the norm group. Both of these results are consistent with the fact that the ESAA students are educationally disadvantaged and would be expected to score lower and gain less than national norm students. The third and most important finding is that the change in CTB scale scores for the period between the end of grade 3 and the beginning

---

*The change in raw score between the end of grade 3 and the beginning of grade 4 is not meaningful since these scores are based on different numbers of items.

Table 6. Raw Score, CTB Scale Score, and ESAA Scale Score Performance for Constant Percentile Performance Across Three Grades: National Norms

Decile	Score Type	National Norms					
		Grade 3		Grade 4		Grade 5	
		September	June	September	June	September	June
	RAW	31	37	20	22	24	28
	CTB	278	300	306	318	329	349
10	ESAA	398.60	437.79	486.19	505.54	524.77	558.01
	RAW	38	47	24.5	28	31	36
	CTB	303	325	331.5	349	363	384
20	ESAA	442.52	479.32	529.16	558.01	580.03	610.31
	RAW	44	58.59	28.67	32.50	36.50	42
	CTB	319	342.50	352.35	370.00	386	407
30	ESAA	467.81	518.27	563.26	589.76	613.13	643.62
	RAW	52.53	66.5	32.33	37.50	41.50	47.5
	CTB	334.5	357.50	369.32	390.00	405	426
40	ESAA	493.07	551.94	588.67	619.00	641.14	672.26
	RAW	62	70.75	36.50	42.00	46.50	53.0
	CTB	348	372.75	386.00	407.00	422.50	444
50	ESAA	531.23	576.93	613.13	643.62	667.24	697.66
	RAW	67.67	73.75	41.00	47.33	51.50	58.0
	CTB	361	388.50	403	425.32	439.50	461
60	ESAA	558.43	598.11	638.66	671.44	690.18	724.70
	RAW	71.50	76.20	46.00	52.5	57	63.0
	CTB	376	405.80	421	442.5	458	480
70	ESAA	582.20	620.35	664.62	695.1	718.55	747.19
	RAW	74.50	78.50	52.5	58.5	62.50	67.5
	CTB	393.50	427.50	442.5	463	478	500.5
80	ESAA	604.20	647.995	695.1	727.17	744.73	772.75
	RAW	77.67	81.00	61.00	65.00	68	72.0
	CTB	419	457	472	389	503	530
90	ESAA	636.55	692.81	738.25	759.36	775.86	804.82

Table 7. Raw Score, CTB Scale Score, and ESAA Scale Score Performance for Constant Percentile Performance Across Three Grades: ESAA Norms

Decile	Score Type	ESAA Norms					
		Grade 3		Grade 4		Grade 5	
		September	June	September	June	September	June
	RAW	27.07	31.16	14.76	18.16	18.56	21.38
	CTB	261.28	278.64	272.32	294.96	297.36	314.28
10	ESAA	366	399.82	422.68	464.62	469.70	499.73
	RAW	32.44	38.83	18.07	21.73	23.01	25.42
	CTB	283.32	305.49	294.42	316.38	323.06	336.1
20	ESAA	409.37	445.76	463.44	503.01	515.17	536.92
	RAW	37.34	47.08	20.78	24.54	26.74	28.99
	CTB	301.02	325.16	310.68	331.70	342.70	353.95
30	ESAA	439.41	479.58	493.94	529.51	547.62	565.76
	RAW	43.18	55.77	23.50	27.41	30.02	33.19
	CTB	316.54	338.77	326.00	346.05	359.08	372.76
40	ESAA	463.86	508.59	519.92	553.12	573.24	594.17
	RAW	50.02	62.53	26.37	30.64	33.74	37.31
	CTB	330.04	349.06	340.85	361.56	374.96	389.24
50	ESAA	489.67	533.28	544.56	577.54	597.63	617.84
	RAW	57.81	68.15	29.60	34.62	38.13	41.93
	CTB	341.62	362.45	357.00	378.48	392.52	406.72
60	ESAA	515.93	561.10	570.20	602.83	622.83	643.27
	RAW	65.26	72.78	33.85	39.40	43.50	46.89
	CTB	354.52	382.68	375.40	397.60	412.50	423.67
70	ESAA	545.34	591.19	598.35	630.20	651.45	669.28
	RAW	70.82	76.37	39.19	45.10	50.36	53.56
	CTB	373.10	407.33	396.76	418.30	435.44	446.24
80	ESAA	577.37	622.10	629.03	659.92	685.00	700.42
	RAW	76.30	79.51	46.97	53.34	58.56	63.30
	CTB	406.70	438.61	423.91	445.36	463.24	481.20
90	ESAA	621.37	664.64	669.55	699.33	727.45	748.99

of grade 4 is consistently negative. For each decile point at the beginning of grade 4, the CTB scale score is smaller than that at the end of grade 3. This finding indicates that even if ESAA evaluation students maintained their decile ranks over time, their gain scores, in terms of CTB scale scores, would be negative. This is exactly the phenomenon found in the impact analyses of the second-year ESAA Evaluation. Thus, because of the location and nature of ESAA students, the CTB scale scores apparently produce an inaccurate picture of academic growth.

By contrast, inspection of the new ESAA scale scores for this same set of data points shows that the change is positive, as would normally be expected. Each decile point has a larger ESAA scale score at the beginning of grade 4 than at the end of grade 3. Furthermore, no negative changes are apparent between any pair of testing points with the new scale scores. It would seem that the rescaling had adjusted the ESAA student performance so that it is consistent with normal expectations of cognitive growth.

To further demonstrate the use of the new scores, the mean reading scores for grade 3/4 longitudinal students in the second-year growth analysis were translated onto this scale. Table 8 contains the original CTB scale scores as well as the ESAA scale score equivalents. The results indicate that a very different achievement pattern is obtained from the ESAA and CTB scale scores, as illustrated in Figures 8 and 9. Figure 8 displays the performance found when the CTB scores were used to represent the growth of the ESAA students. When the same results are reported with the ESAA scale scores (see Figure 9), the achievement pattern is monotonically increasing in all cases. Furthermore, the trend is much more consistent with the assumptions of the Absolute Scaling procedure. It would appear that the anomolous effects of level articulation have been removed from the ESAA evaluation data when the new scale scores are used to represent performances.

It is important to note that while the present study used only the CAT, the major inferences derived from the restandardization effort almost certainly have much wider generalizability. Indeed, a number of studies (Barker and

Table 8. CTB Scale Score and ESAA Scale Score Equivalents of Mean* Reading Performance of Longitudinal Students in Second-Year Evaluation

	Grade Year 1/Year 2	Group	Score Type	Year 1		Year 2	
				Pre	Post	Pre	Post
Pilot	3/4	Treatment	CTB	309	331	327	349
			ESAA	405	490	523	558
		Control	CTB	315	338	338	361
			ESAA	463	506	542	577
Basic	3/4	Treatment	CTB	333	356	358	380
			ESAA	495	549	571	605
		Control	CTB	332	354	359	377
			ESAA	491	544	573	601

*For descriptive purposes, the mean performances have been rounded to the nearest whole integer.

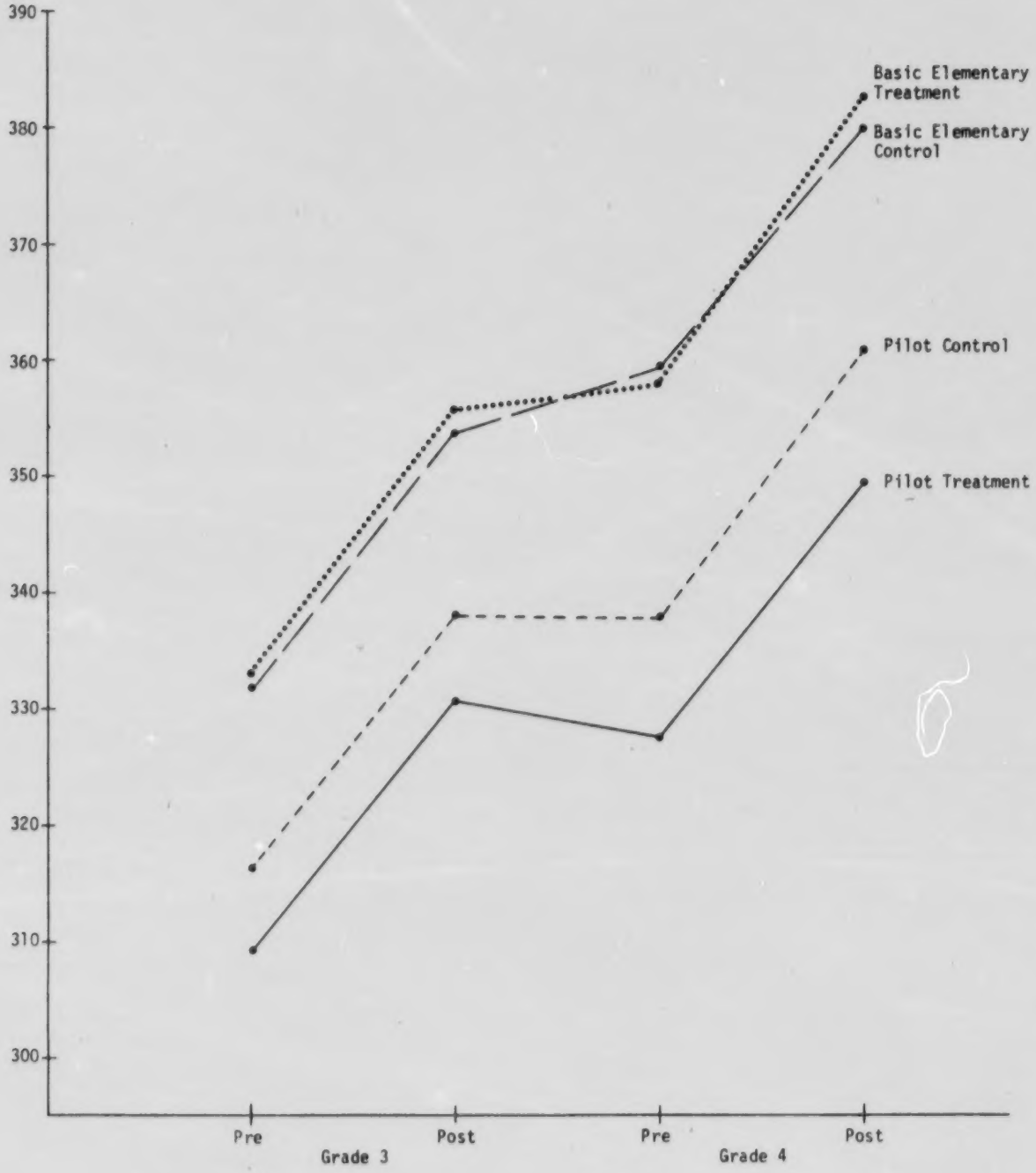


Figure 7. Mean CTB Reading Scale Scores for Pre- and Posttestings in Grades 3 and 4

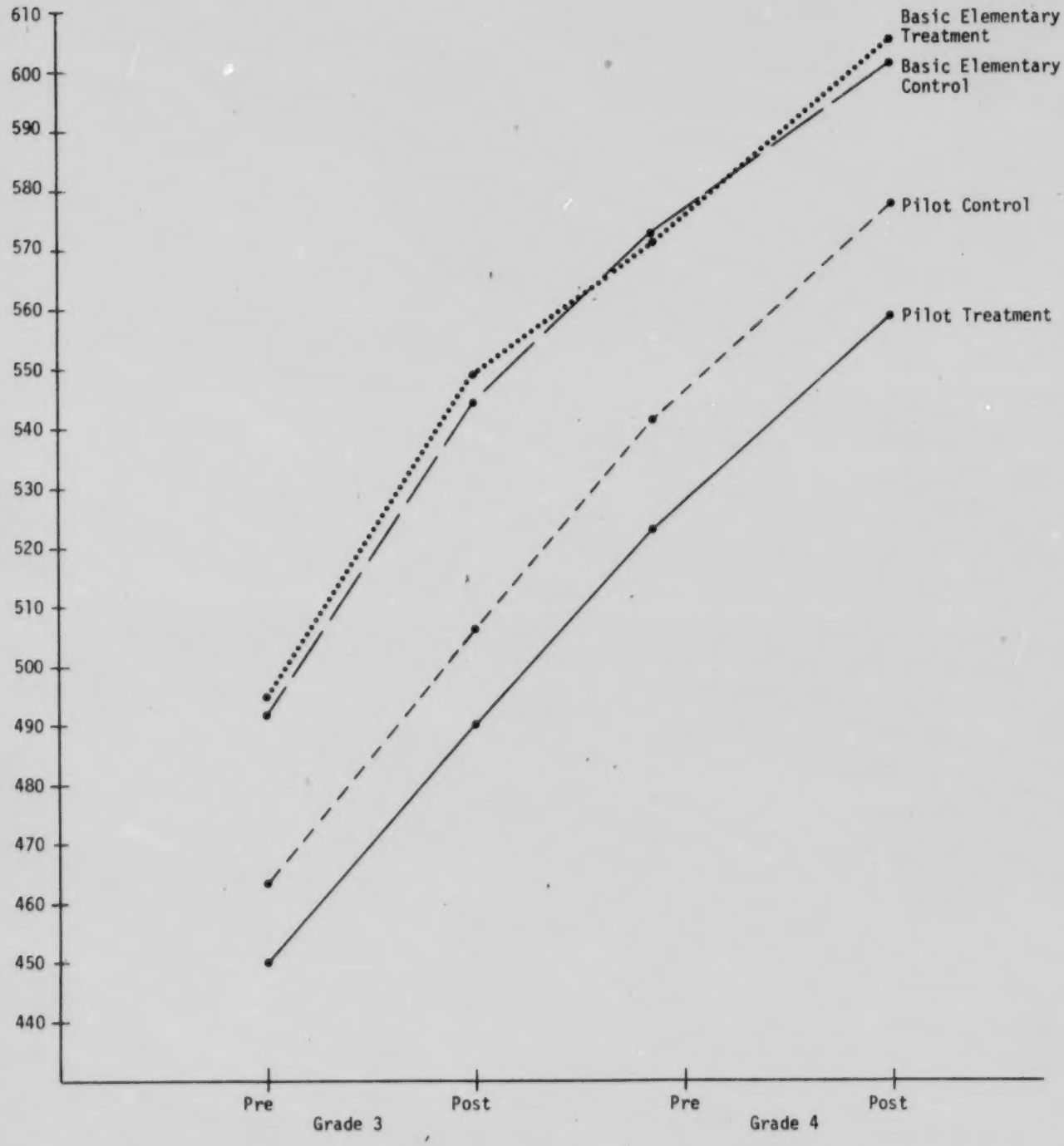


Figure 8. Mean ESAA Reading Scale Scores for Pre- and Posttestings in Grades 3 and 4

Pelavon, 1975; Pelavon and Barker, 1976; Slinde and Linn, 1977; Long, Schaffran, and Kellog, 1977; and Kaskowitz and Norwood, 1977) have recently reported similar findings with other test batteries. These studies have all indicated the existence of level articulation problems, particularly in the case where compensatory education students are concerned.

(3) Results of Equating the Forms

Separate scaling was conducted independently for the two test forms. This was done primarily to ensure that the forms would be comparable in terms of the scores submitted to the analytic techniques used in the ESAA evaluation. The results (see Table 5) show that, after completion of our rescaling activities, the average test results for the two test forms were very similar in terms of the scale scores. However, these were further adjusted by applying the equi-percentile equating technique described in the procedure sections. The mid-percentile rank for each raw score on the Form B test was used as the reference point for the equating procedure. Form A scale scores were calculated which corresponded to these mid-percentile ranks. The final product was a set of scale scores for Form B that were directly comparable to those for Form A. Thus, a person at a specific percentile rank would receive the same scale score regardless of the test form used to measure performance. Tables of equivalent scores are reported in another document (Doherty, 1977).

As mentioned previously, a major problem to be rectified by the restandardization study was the elimination of inadequacies of the form-to-form equating. While the above procedure theoretically should have produced the desired results, its effectiveness was empirically investigated. Tables 9 and 10 present Form B results for the national norm sample and for the ESAA restandardization sample. The raw, CTB scale, and ESAA scale scores have been reported for decile performance within the two samples for six testing points. The same information is reported for Form A in Tables 6 and 7. Comparing the results in Table 6 with those in Table 9, and Table 7 with Table 10, provides information regarding the effectiveness of the form-to-form equating for both sets of norms.

Table 9. Raw Score, CTB Scale Score, and ESAA Scale Score Performance (Form B) for Constant Percentile Performance Across Three Grades: National Norms

National Norms

Decile Rank		Grade 3		Grade 4		Grade 5	
		Fall	Spring	Fall	Spring	Fall	Spring
10	Raw	28.0	33.0	16.0	17.5	19.5	22.5
	CTB	278	298	306	317	330	349
	ESAA	401	437	461	480	508	548
20	Raw	34.0	41.0	20.0	22.3	25.0	29.0
	CTB	302	325	333	348	363	384
	ESAA	445	483	515	546	578	612
30	Raw	39.0	47.0	23.0	26.3	29.3	34.0
	CTB	319	342	352	371	386	407
	ESAA	474	508	555	592	614	642
40	Raw	44.0	54.5	26.0	30.3	33.5	39.0
	CTB	334	358	369	391	405	426
	ESAA	496	539	589	621	640	668
50	Raw	49.5	62.0	29.3	34.0	38.0	45.0
	CTB	348	373	386	407	423	445
	ESAA	518	573	614	642	662	696
60	Raw	56.0	68.0	33.0	38.5	43.5	51.0
	CTB	361	388	403	425	441	462
	ESAA	545	603	637	665	690	717
70	Raw	63.0	72.3	37.5	44.0	49.5	58.0
	CTB	375	406	421	442	458	480
	ESAA	578	635	660	692	711	738
80	Raw	69.0	75.7	44.0	51.0	57.0	64.5
	CTB	392	427	442	462	477	501
	ESAA	609	669	692	717	735	774
90	Raw	74.5	79.0	54.5	61.0	65.0	71.0
	CTB	419	456	471	489	503	533
	ESAA	657	713	729	755	777	808

Table 10. Raw Score, CTB Scale Score, and ESAA Scale Score Performance (Form B) for Constant Percentile Performance Across Three Grades: ESAA Norms

ESAA NORMS

Decile Rank		Grade 3		Grade 4		Grade 5	
		Fall	Spring	Fall	Spring	Fall	Spring
	Raw	24.2	28.9	12.5	15.2	17.0	17.9
	CTB	263	286	280	300	313	320
10	ESAA	370	408	412	450	474	486
	Raw	28.8	34.1	15.8	18.3	20.1	21.3
	CTB	281	302	304.5	322	334	342
20	ESAA	407	445	458	491	516	533
	Raw	33.1	39.7	18.3	20.6	22.4	24.0
	CTB	298	321	322	337	349	358
30	ESAA	437	477	491	523	547	567
	Raw	37.1	45.5	20.3	22.7	24.5	26.7
	CTB	313	338	335	350	360.5	372
40	ESAA	463	501	519	551	572	595
	Raw	42.3	51.9	22.3	24.7	27.1	29.7
	CTB	329	353	348	362	374	387
50	ESAA	489	528	546	575	599	617
	Raw	49.3	57.5	24.1	27.1	30.5	33.8
	CTB	347	364	358	374	391.5	406
60	ESAA	517	551	568	599	622	641
	Raw	55.4	63.1	26.7	30.7	35.2	40.2
	CTB	360	375	372.5	392	412	431
70	ESAA	543	578	595	623	648	675
	Raw	62.6	69.3	32.1	36.5	42.0	48.3
	CTB	374	393	398	417	436	454
80	ESAA	576	611	631	653	683	708
	Raw	70.9	74.4	41.2	46.8	53.5	58.6
	CTB	400	419	433	450	468	482
90	ESAA	622	656	679	703	725	740

The results of the national norms (reported in Table 6 and 9 respectively) show that the CTB scale scores are fairly well equated. The ESAA scores (Tables 7 and 10) show some discrepancy between decile performance on the two test forms at the extremes of the distributions, particularly at the 3rd-grade level. This condition is probably produced by instability resulting from the small number of data cases at these score levels within the ESAA sample.

A better determination of the adequacy of the form-to-form equating comes from comparing the results in Table 7 and 10. Here, it can be seen that the ESAA scale scores show better equating than the publisher's scale scores, but some discrepancies exist in both sets of scale scores, particularly for the lower percentile ranks. Apparently, the equating has not totally removed all the effects of the inter-form difficulty difference. The results do suggest, however, that the form-to-form equating is sufficiently accurate to preclude misleading results in the analyses of the current ESAA evaluation.

g. Summary and Conclusions

The restandardization study was conducted as part of the total effort to evaluate the federally-funded ESAA program. Because of the special characteristics of the recipients of ESAA aid and of the evaluation design it was desirable to restandardize the published achievement test used as the criterion measure at the elementary grades. The specific purposes of the study were: (1) to construct a common metric applicable across test levels; (2) to equate the alternate test forms so that estimates of student performance would be unaffected by this factor; and (3) to provide percentile rank norms for a population of elementary students similar to those in the ESAA programs. Accomplishment of these objectives required the testing of all ESAA students participating at the beginning of the third-year evaluation. Eight unique testing groups, formed by combinations of grade level, test-level, and test form, were designated, and these provided the data necessary to perform the scaling and equating. Thurston's (1925) Absolute Scaling method, and equipercentile equating (Angoff, 1971), were the basic analytic techniques. Three products were produced: (1) percentile norms for the ESAA population on the California Achievement Test (CAT); (2) scale scores independent of test-level; and (3) scale scores independent of test form and level.

Investigation of the usefulness of new scale scores shows that the new scale scores eliminated most of the level-articulation problem present in the publisher's scale scores. Data from the second-year ESAA evaluation as well as hypothetical data based on constant percentile performance showed that the CTB scores led to improbable conclusions about ESAA student performance, and that the new ESAA scale scores largely rectified this problem. The importance of this finding goes beyond the ESAA study to other evaluations of compensatory education, as it was shown that other studies had found articulation problems in other test batteries when used with disadvantaged, low-achieving students.

2. THE DISTRIBUTION OF ACHIEVEMENT SCORES IN A DISADVANTAGED POPULATION

Much of the research and program experimentation in education is directed toward compensatory programs having as their focus the improvement in achievement of educationally disadvantaged students. The evaluation of such improvement typically is based upon achievement measures that were normalized on populations substantially different from the usual research population. Although it is well known that a disadvantaged population does not perform at the achievement level of a national normative population, the precise pattern of underachievement is not nearly so clear. Consequently, there is little information about what constitutes a realistic goal for a disadvantaged population's improvement in achievement within a specified time period.

Brief descriptive summaries of the sample ESAA students' achievement were presented in reports for each of the three evaluation years. The present report documents in greater detail the ESAA students' achievement. Such documentation is intended to characterize more comprehensively the reading and mathematics attainment of the sample students, and to provide reference or baseline data for ESAA and other compensatory program administrators and researchers working with disadvantaged students. The data tables and figures contained in this report, all of which are presented in numerical order following the text, and as excerpts within the text, should help program administrators and researchers answer the following types of questions:

- How does the achievement score distribution of students in some programs of interest compare with the achievement level of disadvantaged students in a reference sample, i.e., in the ESAA sample?

- Have the within-year achievement gains of a given set of students met the gain expectation derived from ESAA students of similar initial ability?

- Over a two or three-year time span, have the students of interest made gains meeting the expectation provided by a reference sample of ESAA students?

(Before one tries to use these tables as reference points for a disadvantaged population, their limitations should be noted. Data are available only for grades 3, 4, 5, 10, 11, and 12. Only levels 2 and 3 of the California Achievement Test (CAT), level-5 CAT mathematics, and level-2 of the Iowa Silent Reading Test (ISRT) were used. These tables are not directly applicable to situations where other tests are used, except for the reading tests included in the Anchor Test Study, which will be discussed later. It should also be noted that in the ESAA sample elementary students did not take the full mathematics test. One specific section of one subscale was omitted (the 15-item "problems" section for the Problems and Concepts subscale); however, in the tables which follow, adjustments were made to equate the ESAA total math percentiles to the total math percentiles reported by the publisher. The tests were administered in November and May, so comparisons are most accurate for tests administered close to those dates.

a. Description of the ESAA Evaluation Sample

The ESAA evaluation was conducted on a sample of schools receiving ESAA funds. For each ESAA school, a control school from the same district which received no ESAA funds but was otherwise comparable to the ESAA-funded school was also selected. Within each school, three grades were selected to be analyzed. These were the 3rd, 4th, and 5th grades at the elementary level, and 10th, 11th, and 12th grades at the secondary level. The student samples were drawn randomly across classrooms within grades.

An analysis of the achievement test score distributions for the 1974-75 and 1975-76 school year evaluations revealed no substantial differences; therefore, the data from these two years were pooled so as to enlarge the data base used in constructing these baseline data, and to increase stability of the results. In certain instances that will be noted, the Basic and Pilot Elementary samples were also pooled in order to provide single elementary and secondary level tables.

Over half the districts included were from HEW regions IV (Southeast) and VI (South Central), which reflected the ESAA Basic and Pilot award universes'

concentration in those regions. However, overall, there was a fairly broad geographical representation, especially at the elementary level.

Approximately 45 percent of the Basic Elementary sample students were members of minority groups, while 35 percent of the Basic Secondary and over 80 percent of the Pilot Elementary students were minority group members. Black students constituted the largest minority group with Hispanic students a distant second in numbers. In fact, only in the Pilot Elementary sample were there an appreciable number of Hispanic students.

All the groups had substantial numbers of disadvantaged students, as indicated by the percentages of students judged eligible for Elementary and Secondary Education Act (ESAA) Title I services. The largest percentage of Title I-eligible students was found in the Pilot Elementary schools (65 percent); the corresponding percentages for the Basic Elementary and Basic Secondary samples were 43 and 23 percent, respectively.

The reading and mathematics achievement test scores averaged around the 20th and 30th percentile, based upon national norms. More data concerning achievement levels will be presented in the tables which follow. A detailed description of the demographic characteristics of the districts and schools from which this sample of students was obtained is available in the three yearly ESAA evaluation reports (Coulson, et al, 1974; Coulson, et al, 1975; Coulson, et al, 1976).

In summary, the schools from which these students were sampled can be characterized as having a large percentage of disadvantaged students, a large percentage of minority enrollment, and low reading and mathematics achievement test scores. In these respects, they can be considered similar to other schools that are eligible to receive ESAA or other compensatory education funds. For schools with these general characteristics, the tables presented here should provide a more meaningful set of comparative benchmarks than the usual norms based on a nationally representative sample. The latter type of comparison can lead to unrealistic expectations concerning achievement gain and thus hide what may actually be a beneficial program.

b. Description of Tables and Figures, and Illustrative Examples of Their Uses

To simplify reading of the text, all tables for this section are grouped at the end of the section, and only excerpts from selected tables are included in the text proper for purposes of illustrating how the tables might be used to obtain different kinds of information about the ESAA students' performance.

(1) Percentile Table Descriptions (General)

Tables 11 through 36 describe fall and spring achievement test scores for the ESAA cross-sectional sample of disadvantaged students at grades 3, 4, 5, 10, 11, and 12. This sample can be used as a reference group for evaluating the achievement status of other disadvantaged samples at a particular point in the school year. To provide the detail necessary for such comparisons, percentile rank equivalents are presented for each of the publisher's percentiles for fall and spring administrations, for each test level and form, and at each grade.

At the elementary level, the California Achievement Tests (CAT) published by CTB/McGraw-Hill were used. Levels 2 and 3 were administered at the appropriate grade levels for both reading and mathematics. One short, 15-item subtest was not used in the mathematics test. This was the Problems section of the Problems and Concepts Mathematics Subscale. However, the publisher provided normative information for the composite that was used, and appropriate adjustments were made to equate the ESAA math composite percentiles to the full scale percentiles reported by the publisher. Consequently, the percentile comparisons reported in the tables represent percentile rankings for the publisher's total mathematics score.

The following hypothetical examples illustrate the use of Tables 11-26:
Hypothetical Illustration 1: A compensatory program administrator has scores from the fall administration of the California Achievement Test (CAT). This program involves only 3rd grades, and a reported reading score at the publisher's 26th percentile is of interest. Table 11 reveals that this corresponds to the 39th percentile in the ESAA sample. For a math score at the publisher's 20th percentile, an ESAA percentile of 35 was found (Table 15). Thus, while the obtained reading and mathematics scores were only at the 26th and 20th percentiles

on the publisher's norms, they are at the 39th and 35th percentiles when compared to a more representative reference group. Excerpts from Tables 11 and 15 appear below:

Excerpt 1: Excerpts from Tables 11 and 15

Table 11 Excerpt. Fall, Total Reading Scores: Grade 3, Test-Level 2

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1	44	60	1	1	48-49	60
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
24	36	72-73	83	26	41	74	84
25	37	74-76	84	27	42	75	85
26	39	77	85	28	43	76	86
27	40	78-79	86	29	44	77	87
28	41	80-82	87	30	45	78-79	88

Table 15 Excerpt. Fall, Total Math Scores: Grade 3, Test-Level 2

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-3	39	58	1	1-4	40	52
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
18	32	62	78	18	29	58	72
19	33	63	79	19	30	59	72
20	35	64-65	80	20	31	60	74
21	36	66-67	81	21	32	61	75
22	37	68	82	22	33	62	76

Hypothetical Illustration 2: In a compensatory program at the secondary level, the following spring reading and mathematics scores are of interest:

10th grade - 21st percentile reading, 28th percentile math
11th grade - 24th percentile reading, 25th percentile math
12th grade - 33rd percentile reading, 31st percentile math

The corresponding percentiles based on the ESAA sample are:

10th grade - 46th reading (Table 31), 65th math (Table 34)
11th grade - 48th reading (Table 32), 61st math (Table 35)
12th grade - 60th reading (Table 33), 70th math (Table 36)

This illustration reveals a situation where scores that were only at roughly the first quartile when compared to the publisher's norm sample actually are at the median or above for the ESAA sample. Excerpts from these tables appear below:

Excerpt 2: Excerpts from Tables 31-36

Table 31 Excerpt. Spring, Total Reading Scores: Grade 10, Test-Level 2

FORM E				FORM F			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-5	40-41	66	1	1-3	40-41	66
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
19	43	67-68	87	19	43	71-72	85
20	44	69-71	88	20	44	73	86
21	46	72	89	21	46	74	87
22	47	73	90	22-23	47	75-76	88
23	48	74-75	91	24	49	77-78	89

Table 32 Excerpt. Spring, Total Reading Scores: Grade 11, Test-Level 2

FORM E				FORM F			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-3	39-40	64	1	1-3	45	68
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
22	46	70-71	86	23-24	46	81-83	90
23	47	72-73	87	25	48	84-85	91
24	48	74-76	88	26	49	86	92
25	49	77	89	27	51	87-88	93
26	50	78-79	90	28-29	52	89-91	94

Table 33 Excerpt. Spring, Total Reading Scores: Grade 12, Test-Level 2

FORM E				FORM F			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-5	40	67	1	1-4	42-43	67
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
30	58	90-91	96	31	59	91-93	97
31	59	92-93	97	32-33	60	94-96	98
32-33	60	94-96	98	34-35	61	97-99	99
34-35	61	97-99	99	36	62		
36	62			37	63		

Table 34 Excerpt. Spring, Total Math Scores: Grade 10, Test-Level 5

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-8	49	80	1	1-7	46-47	77
.	.			.	.		
.	.			.	.		
.	.			.	.		
25	63			25-26	61		
26-27	64			27	63		
28	65			28	64		
29-30	66			29	65		
31	67						

Table 35 Excerpt. Spring, Total Math Scores: Grade 11, Test-Level 5

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-10	46-47	77	1	1-8	44	74
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
23	59	95-96	98	22	57	89-90	95
24	60	97-99	99	23	58	91-92	96
25	61			24	59	93-95	97
26-27	63			25	60	96-97	98
28	64			26-27	61	98-99	99

Table 36 Excerpt. Spring, Total Math Scores: Grade 12, Test-Level 5

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-16	50-51	80	1	1-14	46-47	77
.	.			.	.		
.	.			.	.		
.	.			.	.		
28	68			28-29	65		
29-30	69			30	66		
31	70			31-32	67		
32-33	71			33	68		
34	72			34-35	69		

Hypothetical Illustration 3: A 5th-grade, fall, Form B mathematics score at the publisher's 38th percentile corresponds to an ESAA percentile of 56 (Table 18). A 4th grade, fall, level-2, Form A reading score at the 52nd percentile of the publisher's norms corresponds to the 70 percentile in the ESAA sample (Table 12). A 4th-grade, spring, Form B mathematics score at the 16th percentile of the publisher's norms corresponds to the ESAA 34th percentile (Table 23). Excerpts from Tables 12, 18, and 23 appear below:

Excerpt 3: Excerpts from Tables 12, 18, and 23

Table 12 Excerpt. Fall, Total Reading Scores: Grade 4, Test-Level 2

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-3	40-41	61	1	1-3	42	61
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
7	13	48-49	68	8	15	48-49	69
8	14	50	69	9	17	50	70
9	15-17	51-52	70	10	18	51-52	72
10-11	18	53-54	71	11	20	53	73
12	19	55	72	12	21	54	74

Table 18 Excerpt. Fall, Total Math Scores: Grade 5, Test-Level 3

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-3	40-41	54	1	1-3	40-41	59
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
35	50	87-88	87	36	54	92-94	93
36	51	89-91	88	37	55	95	94
32-38	52	92	89	38	56	96	95
39	53	93	90	39	57	97	96

Table 23 Excerpt. Spring, Total Math Scores: Grade 4, Test-Level 3

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-3	42-43	56	1	1-3	42-43	60
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
14	29	61	69	13	32	60-61	74
15	30	62	70	15	33	62-63	75
16	31	63	71	16	34	64-65	76
17	32	64	72	17	35	66	77
18	34	65-66	73	18-19	37	67-68	78

(2) Separate Percentile Table Descriptions for ESAA Basic and Pilot Samples

Tables 37-52 list the percentiles for the Basic Elementary and Pilot Elementary samples separately. This will allow administrators or researchers to use tables more closely tailored to their particular school characteristics. The user can determine from the sample description section (a), above, whether the characteristics of the school of interest are more closely approximated by the Basic or Pilot schools and then use the appropriate tables to determine the ESAA percentile corresponding to a given publisher's percentile. If it is not clear which to use, it is recommended that the more general tables (11-36) be used.

For Tables 37-52, the title indicates grade and test-levels and whether the table refers to the Basic or Pilot sample. In each case, the first column contains publisher's percentile scores (at five-percent intervals); the second column the Form A, fall ESAA percentile corresponding to that publisher's percentile point; and the third column the Form A, spring ESAA percentile corresponding to the given publisher's percentile. Similar ESAA percentiles for Form B are reported in the last two columns.

As an illustration, suppose a compensatory program administrator is dealing with an elementary school which has a high percentage of disadvantaged students (90 percent), and has yet to implement a program designed to overcome the adverse effects of minority group isolation. This administrator could choose to use the Pilot sample tables since this school is clearly like the Pilot schools in many respects. A 5th-grade, fall, mathematics publisher's score at the 10th percentile would correspond to a Pilot ESAA percentile of 20 (Table 44). A 3rd-grade, spring reading score at the 20th percentile of the publisher's norms would correspond to 42 for the Pilot ESAA sample (Table 37).

As a second illustration, consider the case of a researcher who wishes to evaluate a Basic Elementary ESAA program. The obvious choice for a comparison group would be the Basic Elementary sample. A 4th-grade, spring reading score at the 40th percentile of the publisher's norms would correspond to a Basic ESAA percentile of 59 (Table 47). A 5th-grade, fall mathematics score at the 22nd percentile would correspond to a Basic ESAA percentile of approximately 30 (Table 52). Excerpts from the tables used in these illustrations appear below:

Excerpt 4: Excerpts from Tables 37, 44, 47, and 52

Table 37 Excerpt. Total Reading Scores: Pilot, Grade 3, Test-Level 2

Pub. Percentile	FORM A		FORM B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	12	11	11	12
10	26	23	26	27
15	34	33	36	37
20	42	41	45	45
25	48	48	50	50
30	54	58	56	56

Table 44 Excerpt. Total Math Scores: Pilot, Grade 5, Test-Level 3

Pub. Percentile	FORM A		FORM B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	12	14	15	15
10	20	24	27	30
15	31	31	34	39
20	36	41	45	45

Table 47 Excerpt. Total Reading Scores: Basic, Grade 4, Test-Level 3

Pub. Percentile	FORM A		FORM B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	12	9	10	6
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
30	49	47	48	48
35	54	55	56	56
40	58	59	60	61
45	63	65	67	65

Table 52 Excerpt. Total Math Scores: Basic, Grade 5, Test-Level 3

Pub. Percentile	FORM A		FORM B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	10	10	10	12
10	17	19	19	22
15	26	24	25	29
20	30	32	32	34
25	37	39	38	41
30	41	43	41	47

(3) Longitudinal Gains

While the tables discussed above may provide suitable benchmarks for assessing the performance of similar students at a single testing point, they may not provide the best comparison data for assessing the achievement gains of some group of interest. For this latter purpose, the best benchmark data available from the ESAA evaluation are those based on the study's longitudinal sample, i.e., the subset of students who were tested in the 3rd grade during the first year of the evaluation, and who were also tested in succeeding grade levels in the second and third years of the study.

Figures 9-12 show the profiles of ESAA achievement scores for the longitudinal and Year Three cross-sectional samples for both publisher's (CAT) and ESAA scale scores for the elementary sample. There were too few secondary students in the longitudinal sample to allow a meaningful longitudinal profile to be constructed. The ESAA scale scores are scores developed in an achievement test restandardization using the ESAA sample as the norm group and are discussed in more detail in Doherty (1977). These figures allow an assessment of scale score changes over one or more years of school. For example, the ESAA longitudinal sample showed a gain of approximately 40 ESAA scale points or 20 CAT scale points in reading during the 3rd grade (Figures 9 and 11).

As a further refinement, the longitudinal sample was broken into decades based on the fall testing in Year One. That is, ten groups were formed by breaking

the sample at each decile. The nth decile is the score below which n percent of the sample falls, i.e., ten percent of the sample have scores below the first decile, decade two contained students scoring between the first and second deciles, etc., and decade ten contained students scoring above the ninth decile. These groups were then followed for the six test periods (pretest and posttest) in each of three years. Figures 13-16 display the profiles of reading and mathematics for these groups.

An illustration of the use of these figures follows: Suppose the 3rd grade in some school of interest has average reading and mathematics scores which are at the 23rd and 19th percentiles on the publisher's norms. Tables 11 and 15 indicate these scores are at the 35th and 32nd percentiles based on the ESAA sample. To compare changes in achievement scores in this school to a comparable ESAA reference group, use the profile for the fourth decade (between 30th and 40th percentile in the ESAA sample) of students displayed in Figures 13-16. This group has a 22-point gain in reading and a 27-point gain in mathematics during the 3rd grade, based on publisher's scaled scores (Figures 15-16). For ESAA scaled scores, the gains were 40-points and 77-points (Figures 13 and 14).

As a second illustration, consider the case of a school whose students score at approximately the sixth percentile in reading on publisher's norms. This would place most of the grades in the second group, based on the ESAA sample (Tables 11-18), so the second decade profile on Figures 13-16 would provide a representative comparison group for these schools. This group displayed CAT reading gains of 23, 21, and 7 during the 3rd, 4th, and 5th grades (Figure 15); while the corresponding ESAA scale gains were 40, 31, and 8 (Figure 13).

(4) Percentile Ranks of Gain Scores

Tables listing percentile ranks of gain scores are not generally available, since one usually refers to norm tables in reference to a particular test administration and since tests are not normally administered more than once in a given year. However, in program evaluations, gains are often of greater interest than scores on a particular test administration. To assist in the interpretation of gains, percentile ranks relating to within year scale score gains were calculated.

Figure 9. Profile of ESAA Scale Reading Scores

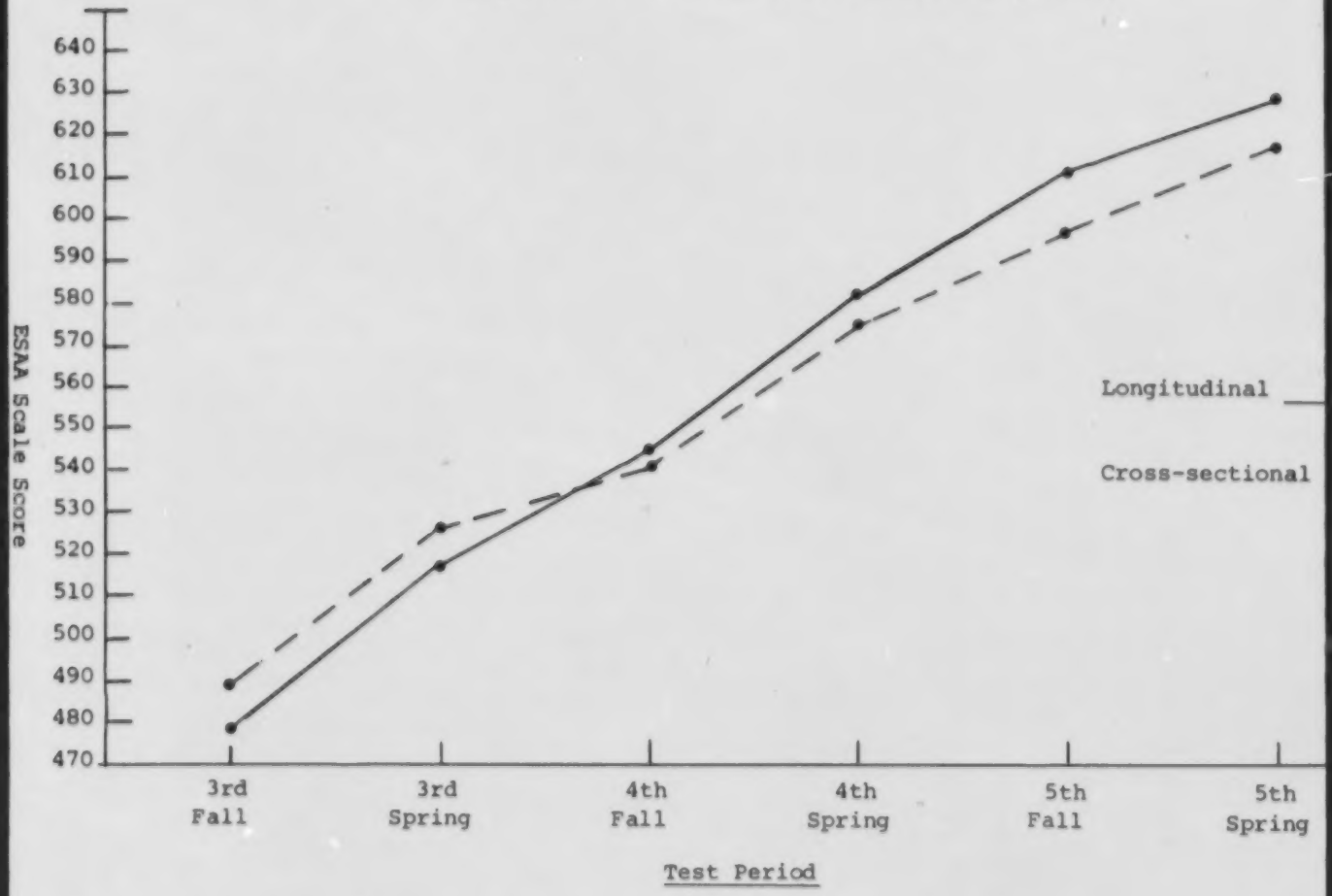


Figure 10. Profile of ESAA Scale Math Scores

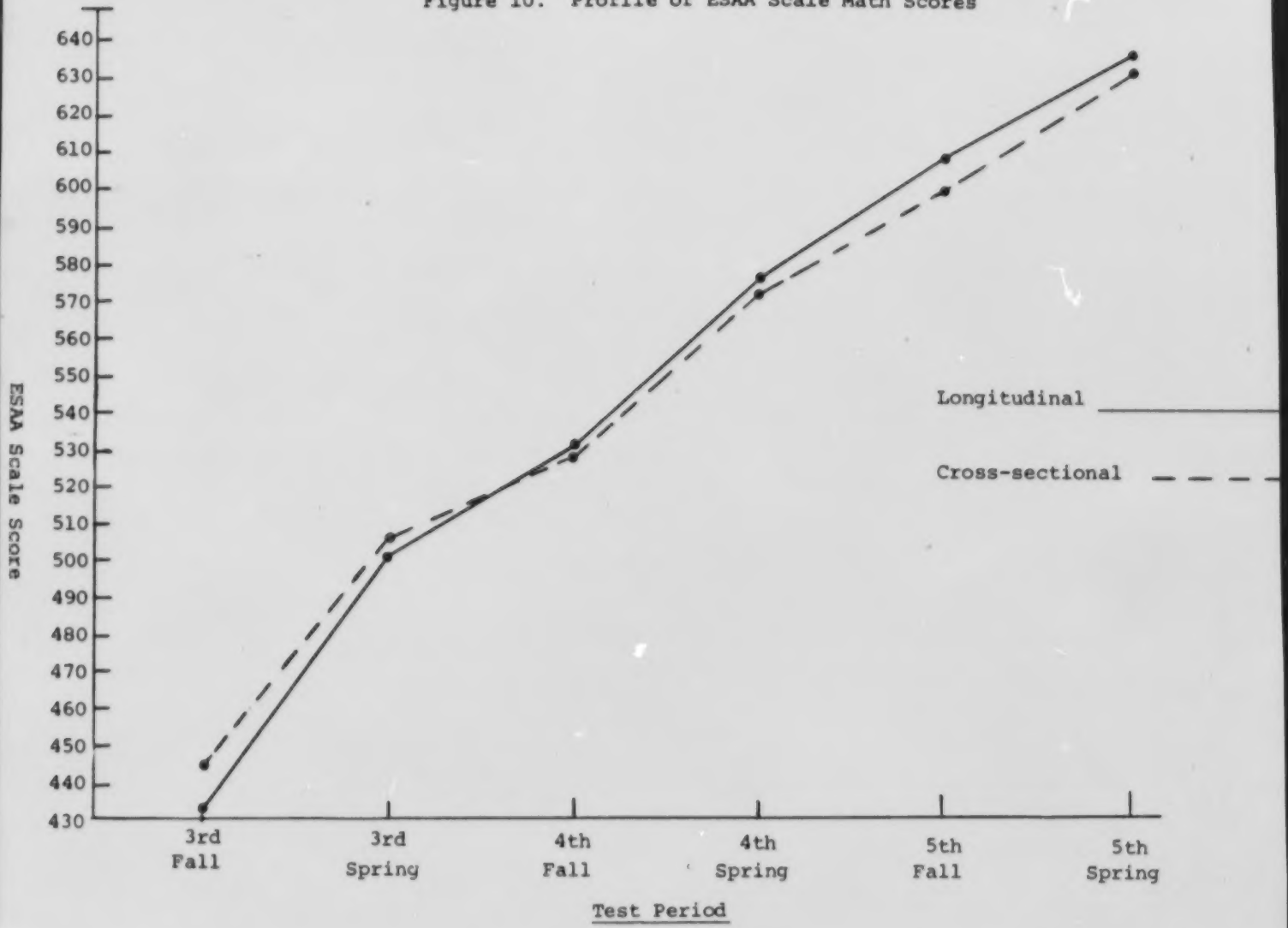


Figure 11. Profile of CAT Scale Reading Scores

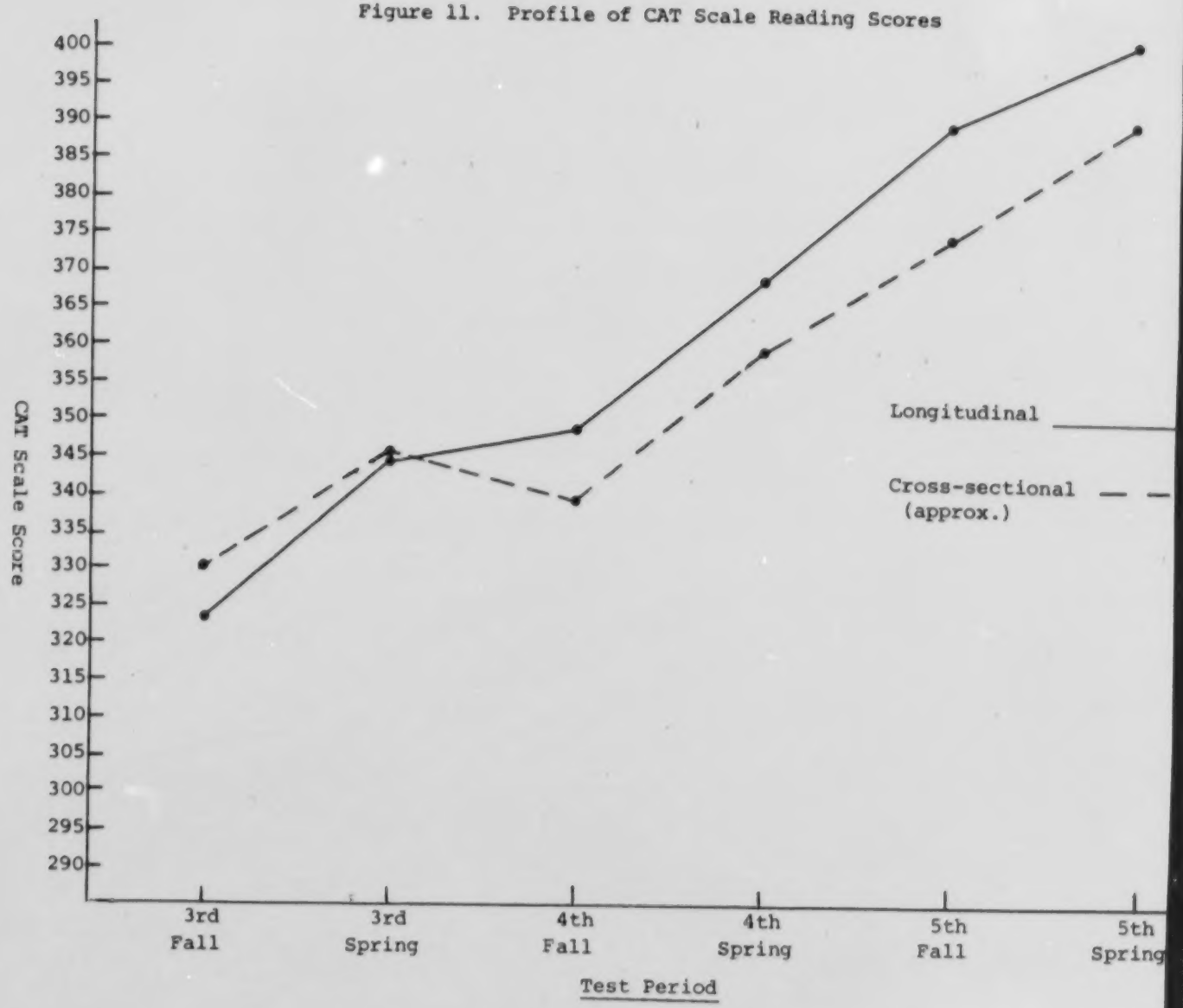


Figure 12. Profile of CAT Scale Math Scores

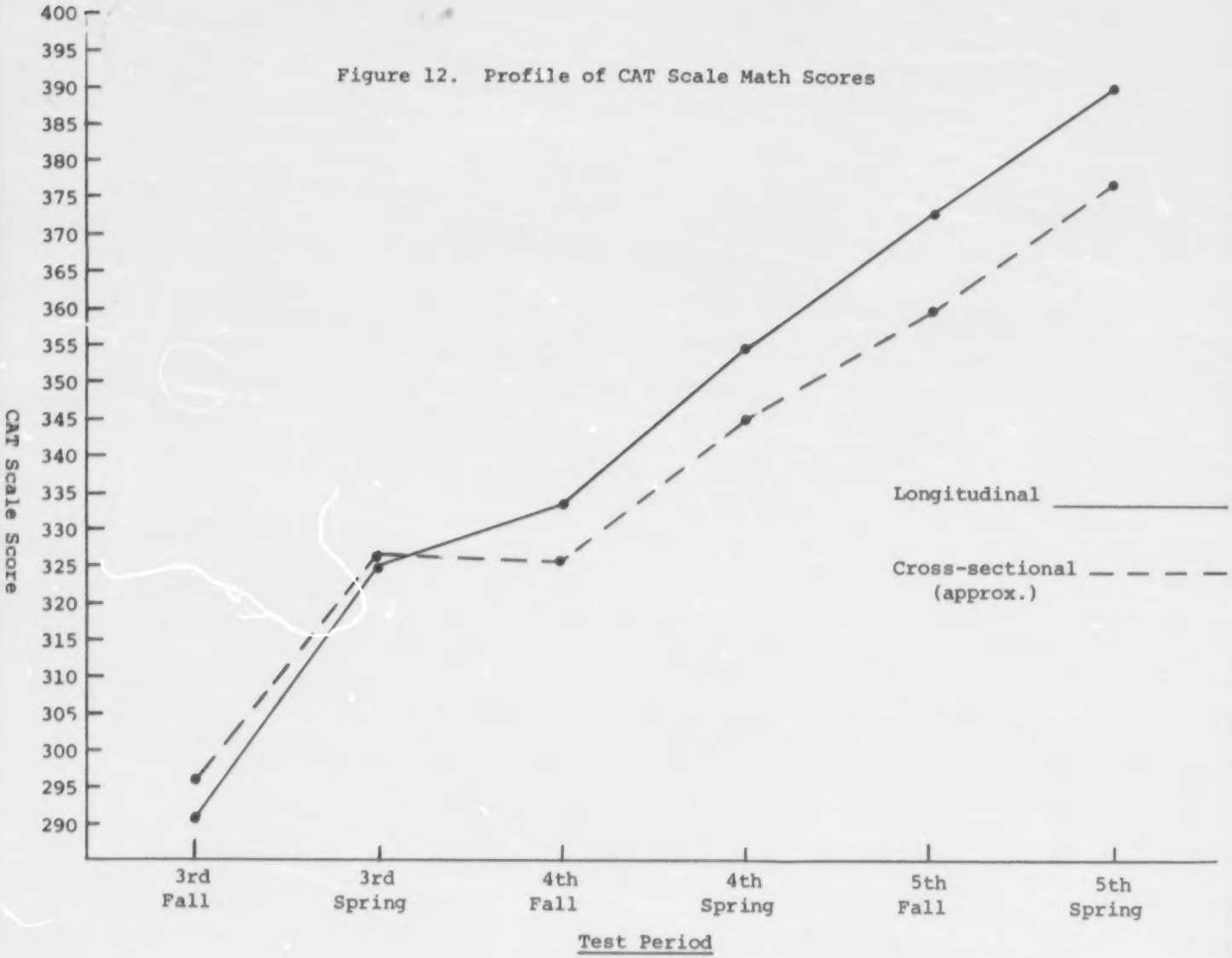


Figure 13. Longitudinal Profile of ESAA Scale Reading Scores for Each Decade (N=1574)

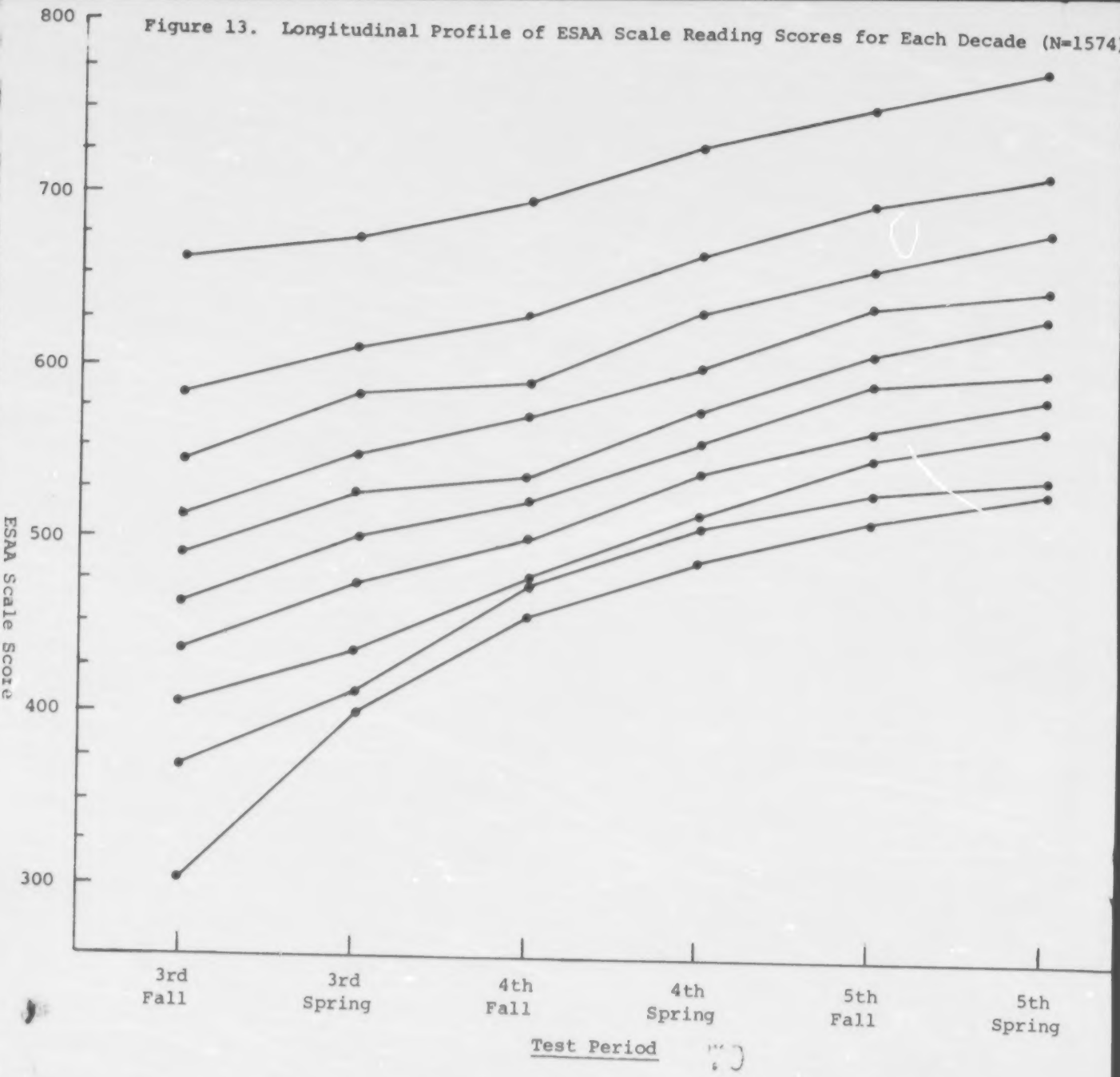


Figure 14. Longitudinal Profile of ESAA Scale Math Scores for Each Decade (N=1574)

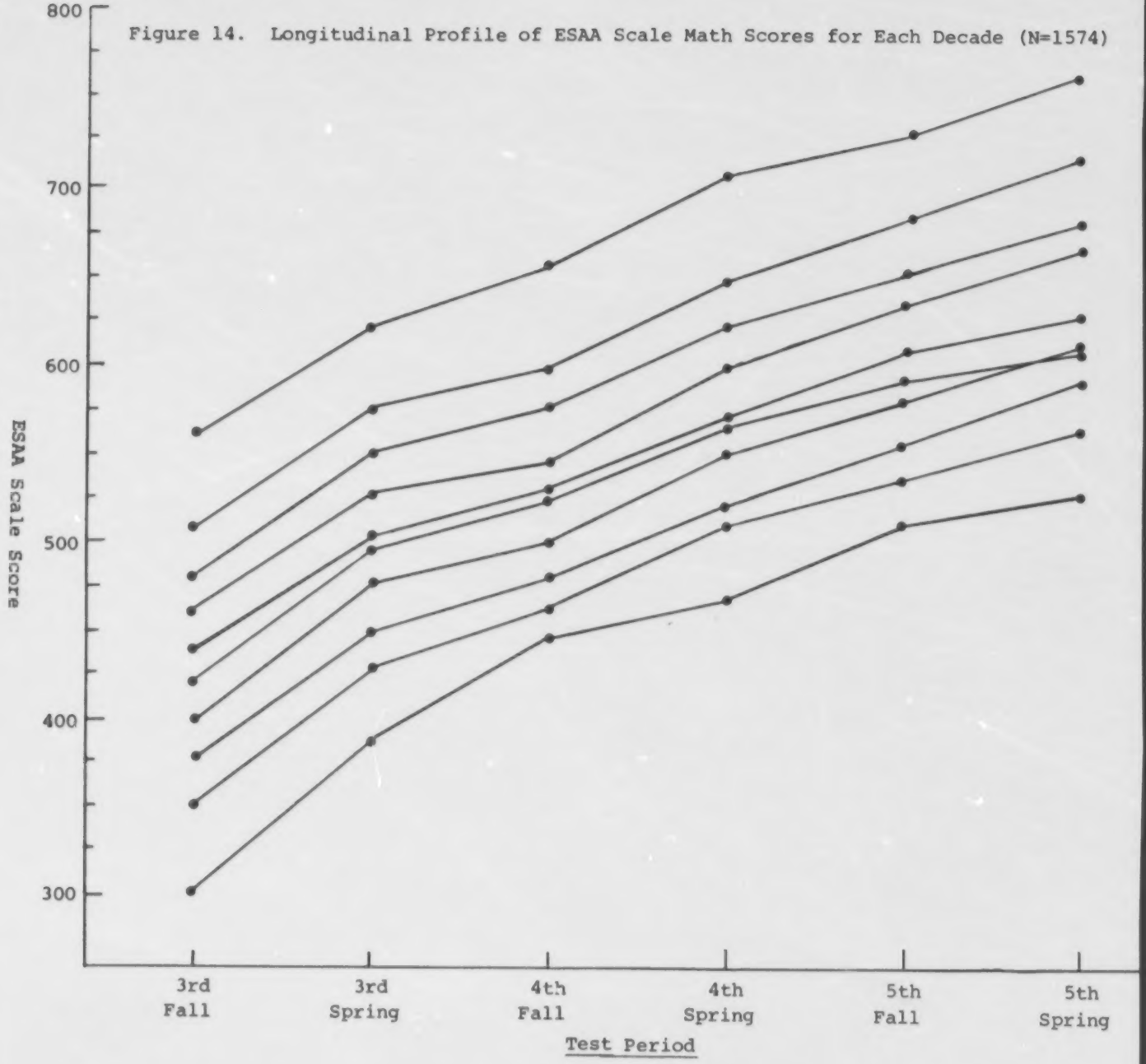


Figure 15. Longitudinal Profile of CAT Scale Reading Scores for Each Decade (N=1574)

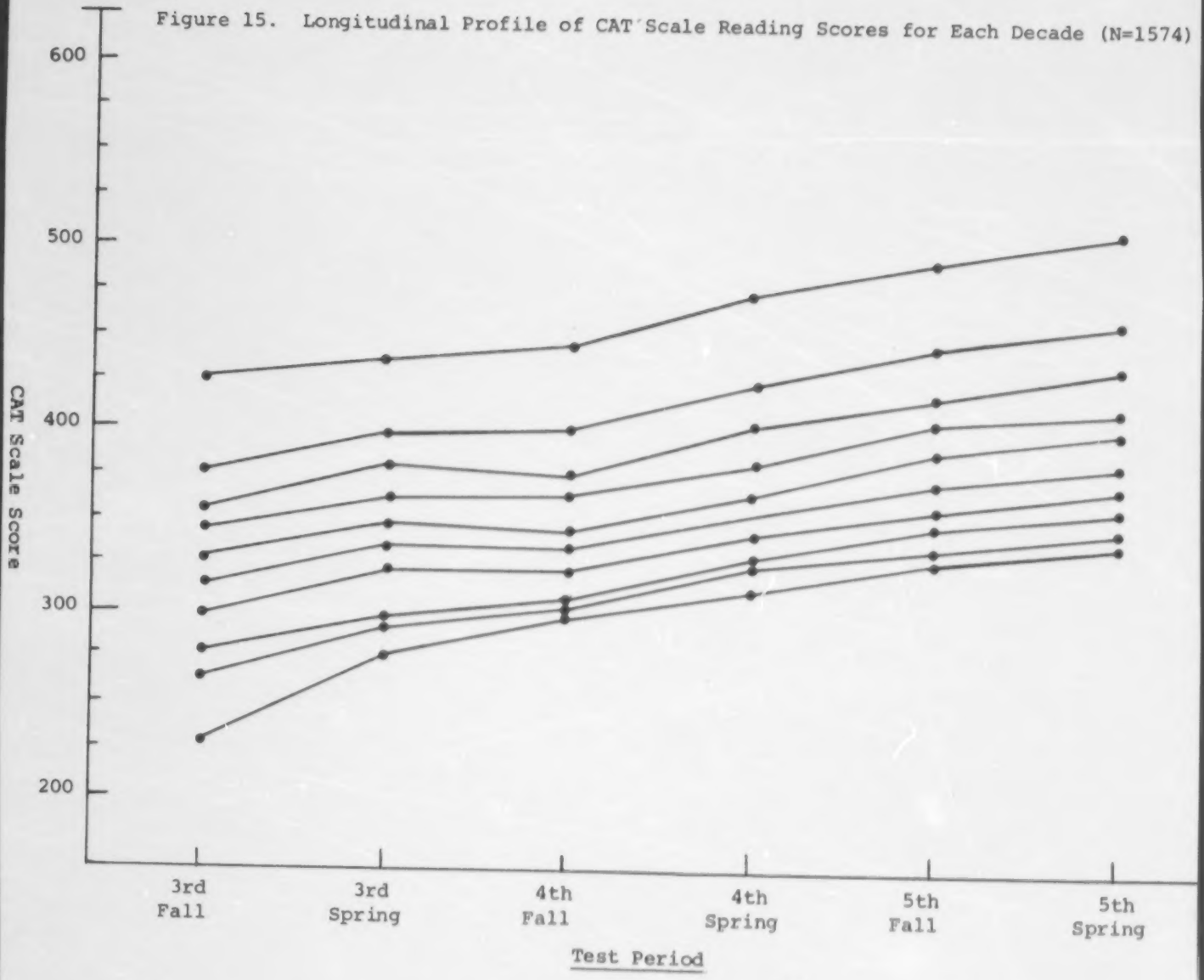
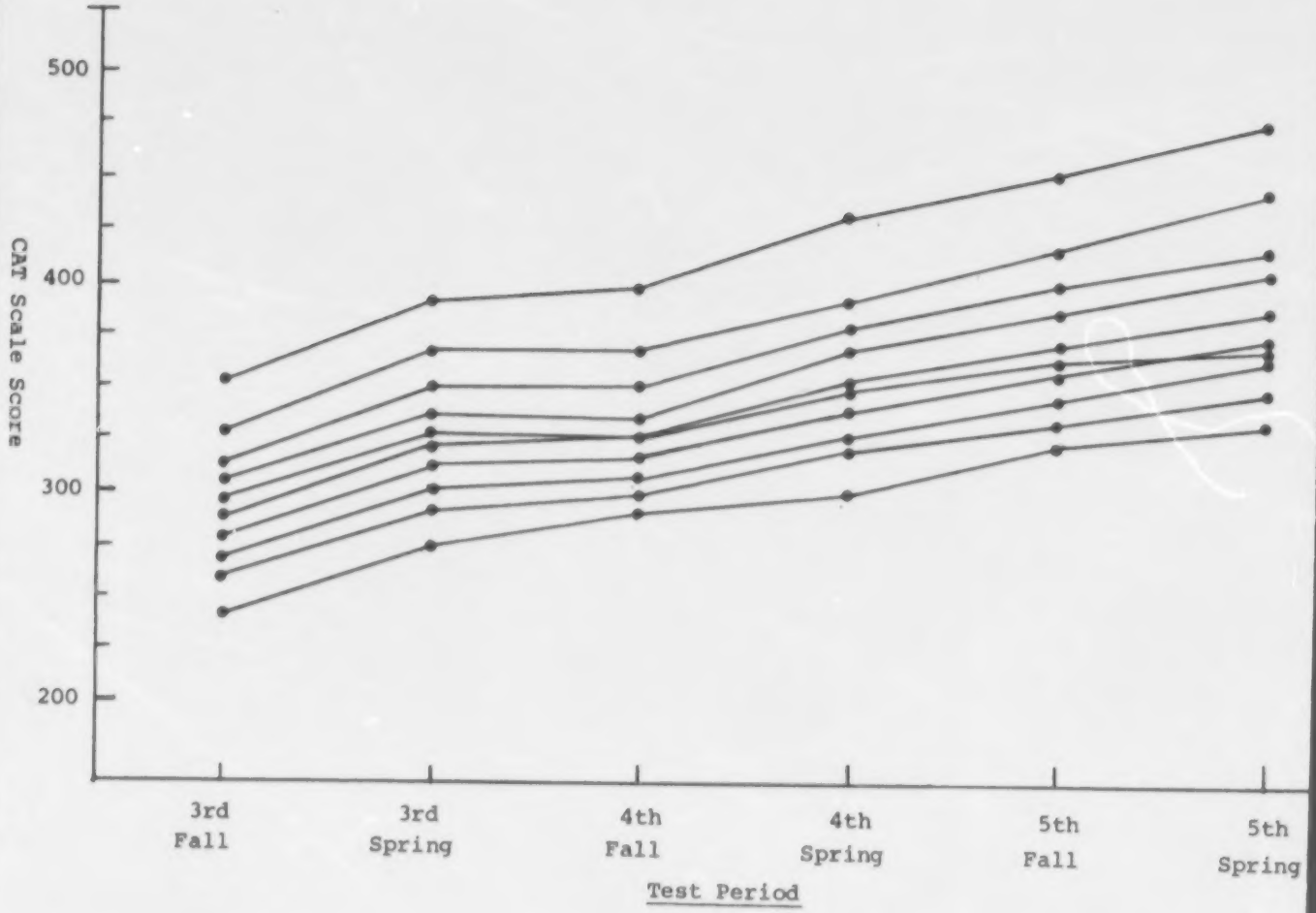


Figure 16. Longitudinal Profile of CAT Scale Math Scores for Each Decade (N=1574)



Tables 53-59 list the percentiles associated with the various gain scores in the ESAA sample. From these tables an educational researcher or a compensatory program administrator can determine how a given gain in achievement compares to the achievement gains registered by a given reference group (the ESAA sample). The table titles indicate the subject area and grade level for each table. The first two columns list ESAA scale gain (in five-point increments) and the percentile associated with that gain. Columns three and four display similar information for the publisher's scale scores. Table 59 contains only percentiles associated with publisher's scale gains for the entire secondary sample since ESAA scale scores were not calculated for secondary students.

As an illustration, suppose a 3rd-grade reading gain of 21 publisher's scale points was observed. Table 53 reveals that this is at the 55th percentile; that is, 55 percent of the ESAA 3rd-grade sample registered a smaller gain. A 5th-grade mathematics gain of five points on the publisher's scale is at the 38th percentile (Table 58). A reading gain of 24 ESAA scale points in the 4th grade is at approximately the 45th percentile (Table 54). A mathematics gain of 11 publisher's scale points at the 12th grade has a percentile rank of approximately 57 (Table 59). That is, within the ESAA 12th-grade sample, 57 percent of the students gained fewer than 12 points in their scaled mathematics score. Excerpts from the tables referenced in this illustration appear below:

Excerpt 5: Excerpts from Tables 53, 58, and 59

Table 53 Excerpt. Distribution of Gain Scores: Reading, Grade 3

ESAA Scale Gain	Percentile Rank	Publisher's Scale Gain	Percentile Rank
No Gain	. 22	No Gain	24
5	25	5	32
10	28	10	32
15	32	15	33
20	36	20	55
25	40	25	56
30	44	40	56

Table 58 Excerpt. Distribution of Gain Scores: Math, Grade 5

ESAA Scale Gain	Percentile Rank	Publisher's Scale Gain	Percentile Rank
No Gain	28	No Gain	30
5	32	5	38
10	35	12	38
15	38	15	45

Table 59 Excerpt. Distribution of Gain Scores: Secondary Sample

Publisher's Scale Gain		Grade 12	
		Reading Percentile Rank	Math Percentile Rank
No Gain		53	47
5		68	52
10		81	57
15		88	62

(5) Anchor Test Study Comparisons

When students take different achievement tests it is not possible to directly compare scores. To alleviate this problem, the Office of Education awarded a contract to Educational Testing Service to conduct an Anchor Test Study. In this study, one reading test (called the Anchor Test) was selected as a common base or anchor, and conversion tables were developed for equating scores on other selected tests to the Anchor Test. Since the California Achievement Test (CAT) was included in that study, it is also possible to link percentiles based on the ESAA norms for the CAT to the various reading tests included in the Anchor Test Study. Tables 60-61 provide these linkages for total reading scores in grades 4 and 5. The Anchor Test Study involved only grades 4, 5, and 6, so it overlapped the ESAA sample only on grades 4 and 5. Values which could not be obtained directly from the Anchor Test Study tables were arrived at by interpolation.

The tests included in Tables 60-61 are the Anchor Test (ANCH), the California Achievement Test--using both the publisher's (CAT) and ESAA norms (ESAA)--the Comprehensive Tests of Basic Skills (CTBS), the Gates-MacGinitie Reading Tests (GMT), the Iowa Tests of Basic Skills (ITBS), the Metropolitan Achievement Tests (MAT), the Stanford Achievement Tests (SAT), the SRA Achievement Series (SRA), and the Sequential Tests of Educational Progress (STEP). The first column of the table lists ESAA percentile points with the corresponding percentile points on the other tests listed in the remaining columns.

As an illustration of the use of these tables, suppose a school has administered the Metropolitan Achievement Test (MAT) to its students. A reported score at the 10th percentile in the 4th grade would correspond to the 35th percentile in the ESAA sample (Table 60). A 5th-grade score at the 19th percentile on the MAT norms would correspond to the 45th percentile in the ESAA sample (Table 61). Scores on other tests can be linked to the ESAA sample in a similar manner. Excerpts from Tables 60-61 appear below:

Excerpt 6: Excerpts from Tables 60-61

Table 60 Excerpt. Approximate Comparison of Selected ESAA Percentiles with Anchor Test Percentiles and Corresponding Publisher's Percentile Ranks on Total Reading Scores: Grade 4

ESAA	ANCH	CAT	CTBS	GMT	ITBS	MAT	SAT	SRA	STEP
99	95	96	95	94	94	93	95	95	95
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
45	20	21	23	22	20	16	14	18	17
40	17	18	19	18	16	13	13	15	14
35	13	15	15	13	12	10	11	12	11
30	10	12	12	8	8	7	9	10	8
25	8	10	11	7	7	6	8	8	6
20	6	7	8	6	6	4	6	6	5

Table 61 Excerpt. Approximate Comparison of Selected ESAA Percentiles and Corresponding Publisher's Percentile Ranks on Total Reading Scores: Grade 5

ESAA	ANCH	CAT	CTBS	GMT	ITBS	MAT	SAT	SRA	STEP
99	98	98	98	97	97	98	96	97	97
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
55	26	28	27	24	21	26	17	24	23
50	23	25	24	21	18	22	15	21	20
45	20	22	21	19	16	18	13	19	17
40	18	19	18	16	14	16	11	16	14
35	15	16	16	13	12	14	10	14	12

c. Additional Illustrative Uses of Test Tables and Figures in this Document

As the illustrations above indicate, the tables presented in this document can be used in various ways to compare other groups of students with the ESAA sample. These tables can be used to make comparisons based on single administrations, to make comparisons of gains (more than one administration), or to chart academic progress. The use of these tables for each of the above purposes will now be discussed in more detail.

(1) Single Administration Comparisons

Step 1. Decide whether to use the Basic or Pilot subsamples as the appropriate comparison group.

Section (a), above, describes some of the differences between the Basic and Pilot Grant schools. If the school (or class) which is of interest seems to be like the Basic Grant schools in the sample, use Tables 45-52. If the Pilot sample more closely approximates the condition, use Tables 37-44. If more detail is wanted, or if it is not clear which of these subsamples more closely approximates the school of interest, use the general Tables 11-36.

Step 2. Decide which test to use for the comparison.

If the California Achievement Test (CAT) was administered at the elementary level, or the Iowa Silent Reading Test at the secondary level, the tables can be used directly. Otherwise, one must check whether one of the reading tests administered in the Anchor Test Study was used. If so, approximate reading percentiles are available in Tables 60 and 61 for grades 4 and 5. No comparable mathematics tables are available.

Step 3. Choose the proper grade level.

Separate tables are provided only for grades 3, 4, and 5 at the elementary level, and grades 10, 11, and 12 at the secondary level. No direct comparisons with the ESAA samples are possible for other grade levels.

Step 4. Choose appropriate test time.

The ESAA sample had tests administered in both the fall and spring of the school year. If the single administration for which the comparison is being made was conducted in the first half of the school year, compare with the columns listing the fall test period data; otherwise, use the columns for the spring test period. The comparisons are most accurate for tests administered in November and May.

Step 5. Select test level.

Comparisons should be made so that the scores of interest were obtained on the same grade at the same level of the test. Direct comparisons are possible (Tables 11-52) for grade 3, level-2; grade 4, level-2 (fall only); grade 4, level-3; grade 5, level-3; grade 10, CAT level-5; grade 10, ISRT level-2; grade 11, CAT level-5; grade 11, ISRT level-2; grade 12, CAT level-5; and grade 12, ISRT level-2.

Step 6. Determine the appropriate form.

Unless it is known that Form B was administered, use the columns of tables referring to Form A.

The following paragraphs explain the use of the tables for single administration comparisons:

A compensatory program administrator is interested in comparing the scores for a given school with the scores from the students in the ESAA sample. This school has about a 60-percent minority enrollment and about 55 percent of the school's students are eligible to receive Title I funds. Since these figures place the school somewhere between the averages for the Basic and Pilot schools in the ESAA sample, the general (pooled) tables will be used for making comparisons. The school is an elementary school where the California Achievement Tests (CAT) were administered about a month and a half before the end of the school year. Level-2 was administered to grade 3 and level-3 was administered to grades 4 and 5. The test form was not noted, so Form A tables will be used. Hence, the tables of interest to this administrator are Tables 19-24. The average percentile scores for each grade, as reported by the publisher, were as follows:

Grade 3 Reading - 13th percentile
Grade 3 Math - 9th percentile
Grade 4 Reading - 11th percentile
Grade 4 Math - 11th percentile
Grade 5 Reading - 14th percentile
Grade 5 Math - 13th percentile

Based on the ESAA sample, rather than the publisher's normative sample, the percentiles for each grade are:

Grade 3 Reading - 22nd percentile (Table 19)
Grade 3 Math - 16th percentile (Table 22)
Grade 4 Reading - 23rd percentile (Table 20)
Grade 4 Math - 22nd percentile (Table 23)
Grade 5 Reading - 35th percentile (Table 21)
Grade 5 Math - 27th percentile (Table 24)

(2) Gain Comparisons

Teachers and administrators are often interested in assessing change over a given school year. Tables 53-59 were constructed to provide a means of comparing the gains attained by a given set of students to the gains exhibited by the ESAA sample. Information is available only for the overall ESAA sample in grades 3, 4, 5, 10, 11, and 12. Comparisons are possible only for scale scores on the California Achievement Test (CAT), and the Iowa Silent Reading Test at the secondary level. At the elementary level, gains based on ESAA scale scores, which were derived from a restandardization of the CAT, are also available; however, these will be of little use unless a copy of the ESAA restandardization of the CAT is also available.

As an illustration, suppose a school district has initiated a new compensatory education program at the elementary level and wishes to evaluate its effectiveness. As part of the evaluation the CAT is administered to all of its students in both the fall and spring. The gains during this school year can then be compared to the gains registered by the ESAA sample. Note that, at the secondary level, only mathematics scores can be considered, since the ISRT was not administered. The district finds the following distributions at the indicated grade levels:

Grade 3 Reading:	No gain - 12% of the students
	Less than 26 points - 50% of the students
	More than 50 points - 20% of the students
Grade 3 Math:	No gain - 8% of the students
	Less than 26 points - 35% of the students
	More than 50 points - 25% of the students
Grade 4 Reading:	No gain - 16% of the students
	Less than 26 points - 50% of the students
	More than 50 points - 18% of the students
Grade 4 Math:	No gain - 20% of the students
	Less than 26 points - 48% of the students
	More than 50 points - 20% of the students

Grade 5 Reading:	No gain - 25% of the students
	Less than 26 points - 53% of the students
	More than 50 points - 21% of the students
Grade 5 Math:	No gain - 26% of the students
	Less than 26 points - 50% of the students
	More than 50 points - 18% of the students
Grade 10 Math:	No gain - 43% of the students
	Less than 26 points - 65% of the students
	More than 50 points - 18% of the students
Grade 11 Math:	No gain - 48% of the students
	Less than 26 points - 69% of the students
	More than 50 points - 12% of the students
Grade 12 Math:	No gain - 51% of the students
	Less than 26 points - 70% of the students
	More than 50 points - 12% of the students

Considered in isolation these data would have little meaning, although the large percentages of students registering no gain indicate an area of possible concern. Is this situation normal? Is the compensatory program having no effect? Tables 53-59 offer added data to help in answering these questions. For the ESAA sample the corresponding percentages were as follows:

Grade 3 Reading:	No gain - 24% of the students
(Table 53)	Less than 26 points - 56% of the students
	More than 50 points - 12% of the students
Grade 3 Math:	No gain - 15% of the students
(Table 56)	Less than 26 points - 42% of the students
	More than 50 points - 21% of the students
Grade 4 Reading:	No gain - 32% of the students
(Table 54)	Less than 26 points - 56% of the students
	More than 50 points - 17% of the students

Grade 4 Math:	No gain - 26% of the students
(Table 57)	Less than 26 points - 53% of the students
	More than 50 points - 15% of the students
Grade 5 Reading:	No gain - 33% of the students
(Table 55)	Less than 26 points - 60% of the students
	More than 50 points - 15% of the students
Grade 5 Math:	No gain - 30% of the students
(Table 58)	Less than 26 points - 57% of the students
	More than 50 points - 14% of the students
Grade 10 Math:	No gain - 39% of the students
(Table 59)	Less than 26 points - 60% of the students
	More than 50 points - 25% of the students
Grade 11 Math:	No gain - 44% of the students
(Table 59)	Less than 26 points - 65% of the students
	More than 50 points - 20% of the students
Grade 12 Math:	No gain - 47% of the students
(Table 59)	Less than 26 points - 68% of the students
	More than 50 points - 19% of the students

This shows that the district's elementary students were performing consistently better than the ESAA sample. At the same time, the secondary students were exhibiting less gains than the ESAA sample, lending support to a claim of beneficial impact of the compensatory program at the elementary level. Note that the selection of points (0, 25, 50 percent) in this example was entirely arbitrary, and the user may choose other points on which to make the comparisons.

(3) Charting Academic Progress

Figures 9-16 can be used to assist in charting the academic progress of an individual student, a classroom, a school, or a district in relation to the progress shown by the ESAA sample. Profiles of both CAT and ESAA scale scores are provided; however, the latter can be used only after reference to the tables provided in the ESAA restandardization report. Profiles are provided only for

grades 3, 4, and 5, since no other elementary grades were sampled and the longitudinal sample at the secondary level was too small to provide a stable profile. Examples will now be presented to illustrate the use of Figures 9-12 and the use of Figures 13-16:

Example 1: A school wishes to compare its achievement profile (in reading) with the profile for the ESAA sample. The CAT is administered in the spring and the average reading scores are as follows: grade 3=349, grade 4=356, grade 5=396. To compare these scores with those for the ESAA sample, first copy Figure 11 and then plot the school's scores on this copy (stars). The resulting figure should be like that in Exhibit 1. The 'profile' provided by these points should be compared with the dashed profile in Figure 11, since these points represent scores for different groups at the same time (a cross-sectional sample). Exhibit 1 shows that this school's average achievement levels are close, but consistently higher than the corresponding points for the ESAA sample.

Example 2: Another school institutes a new mathematics program at the elementary level. The school wishes to evaluate the effectives of this program and decides to administer the CAT to all of its students at the beginning and end of the school year. These exams were not given previously, so no comparison with previous achievement within the school is possible. It is also felt that the program's effects may be cumulative, with greater evidence of impact being associated with more years' exposure to the program. In this case, the school is interested in charting the progress of a fixed group of students over an extended period of time (a longitudinal sample). Figure 12 can be used to provide benchmark data for assessing this progress. To illustrate, consider the 3rd graders involved in the first year of the program. Their scores over the first three years of the program were:

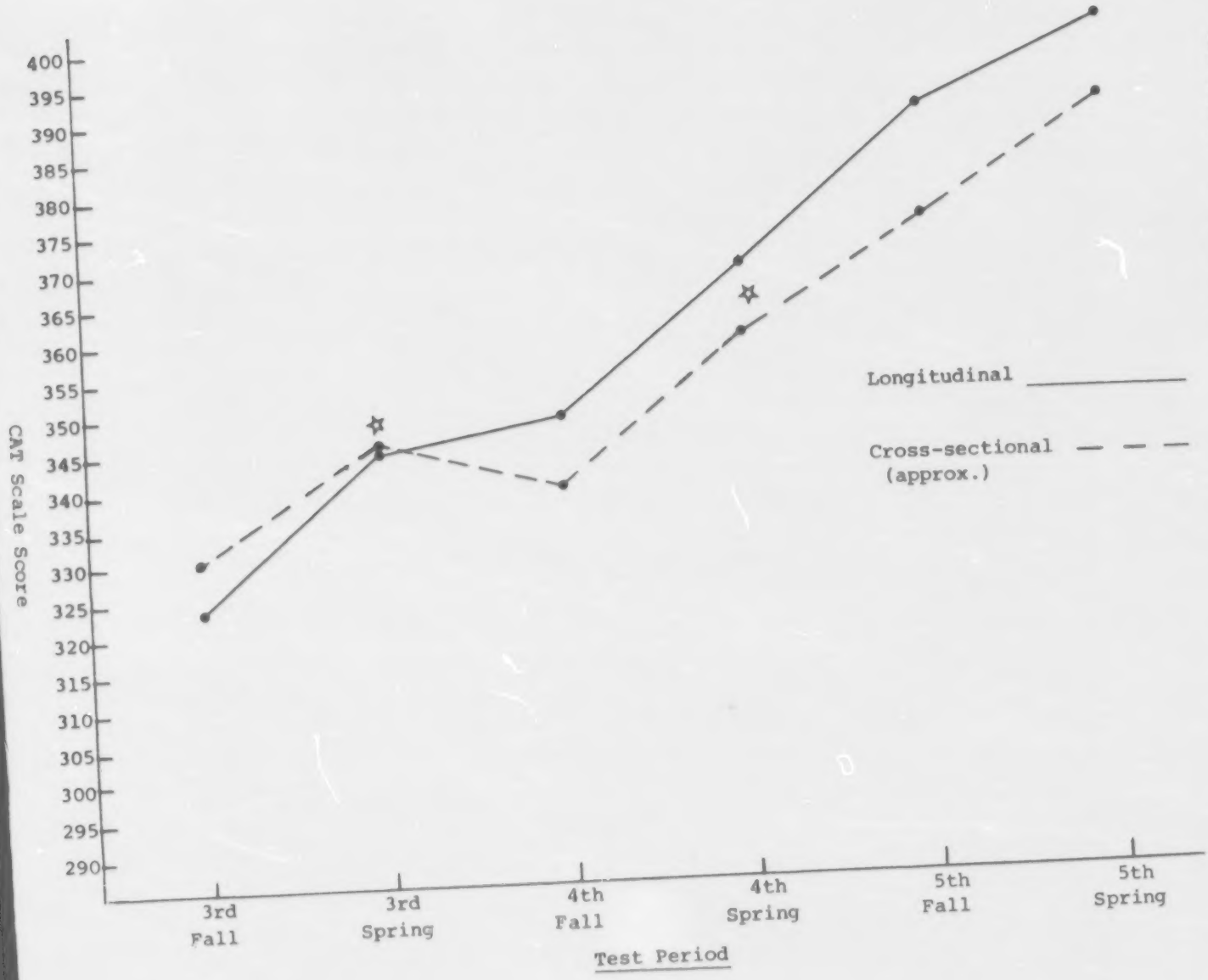


Exhibit 1. The Use of Figures 9-12 to Make Longitudinal Comparisons

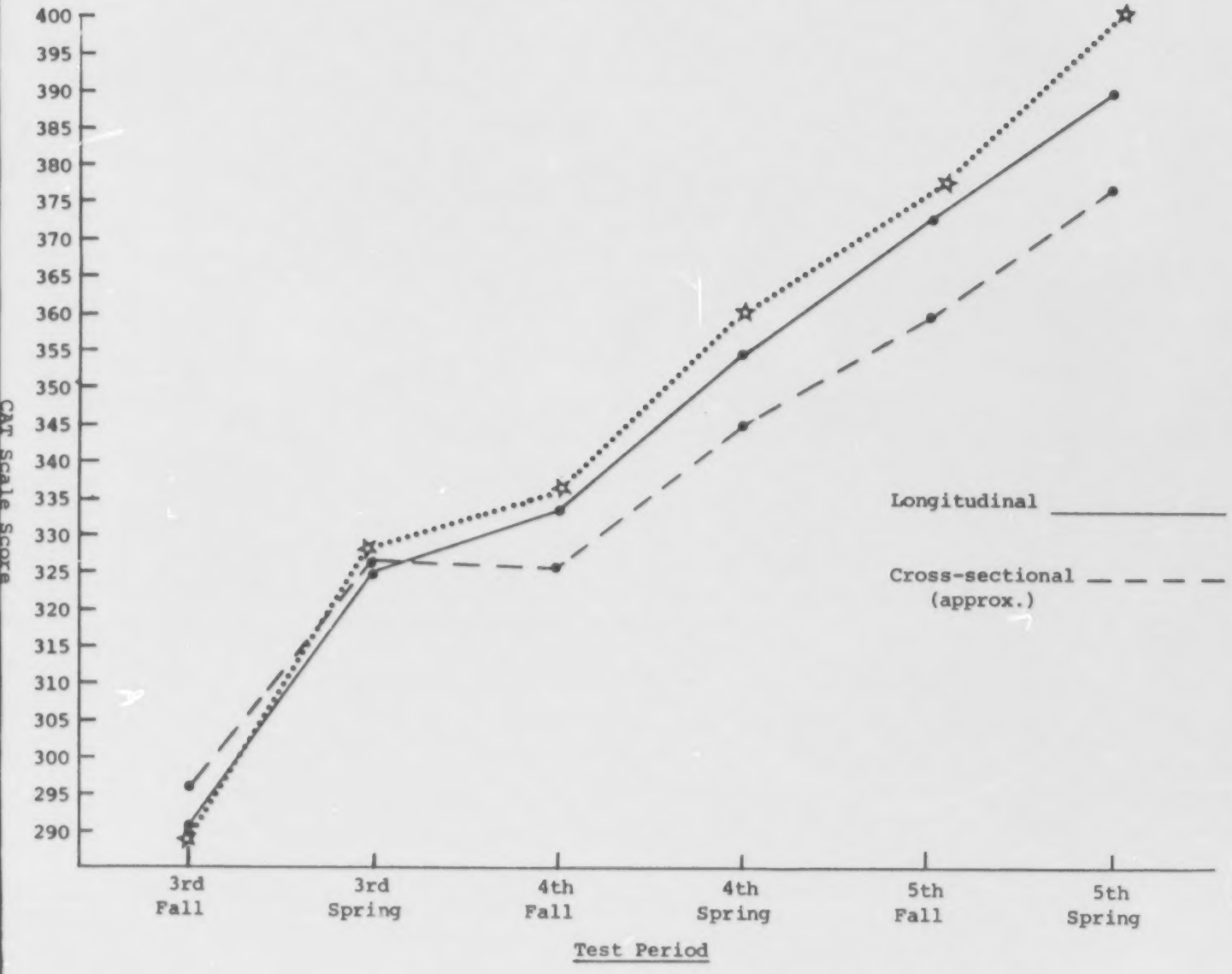


Exhibit 2. The Use of Figures 9-12 to Make Longitudinal Comparisons

	Fall	Spring
Grade 3	289	329
Grade 4	335	360
Grade 5	376	402

To chart these, (a) copy Figure 4 and (b) plot these points (stars) on the copy. The resulting figure should be like that in Exhibit 2 with the dotted profile belonging to this particular school. In this case there is some evidence that the program may, indeed, have had an effect. The students start at a point slightly below that of the ESAA longitudinal sample (the solid profile), but their profile crosses during the first year and the two profiles appear to be diverging over time, beneficially.

Example 3: A school may be interested in obtaining a more refined estimate of its students' academic progress by making comparisons to subgroups within the ESAA sample whose academic standing is more closely related to that of the school. Although there are numerous students who transfer in and out of a school, there is also a great deal of continuity, so comparisons between the school and the ESAA sample could reasonably be made using Figures 15 and 16. Suppose that CAT reading scores are obtained twice a year in a given school. If this is done for several years, the progress of each grade could be plotted on Figure 15. First, make a copy of Figure 15 and then plot the obtained points on the copy. To illustrate, suppose the 3rd grade in a given year is followed over a three year interval. The reading scores obtained are:

	Fall	Spring
Grade 3	280	301
Grade 4	325	347
Grade 5	376	393

The resulting figure should be like that in Exhibit 3, with the dashed line representing the school's profile. This shows that the school has substantially improved itself, in relation to the ESAA sample over this three-year period.

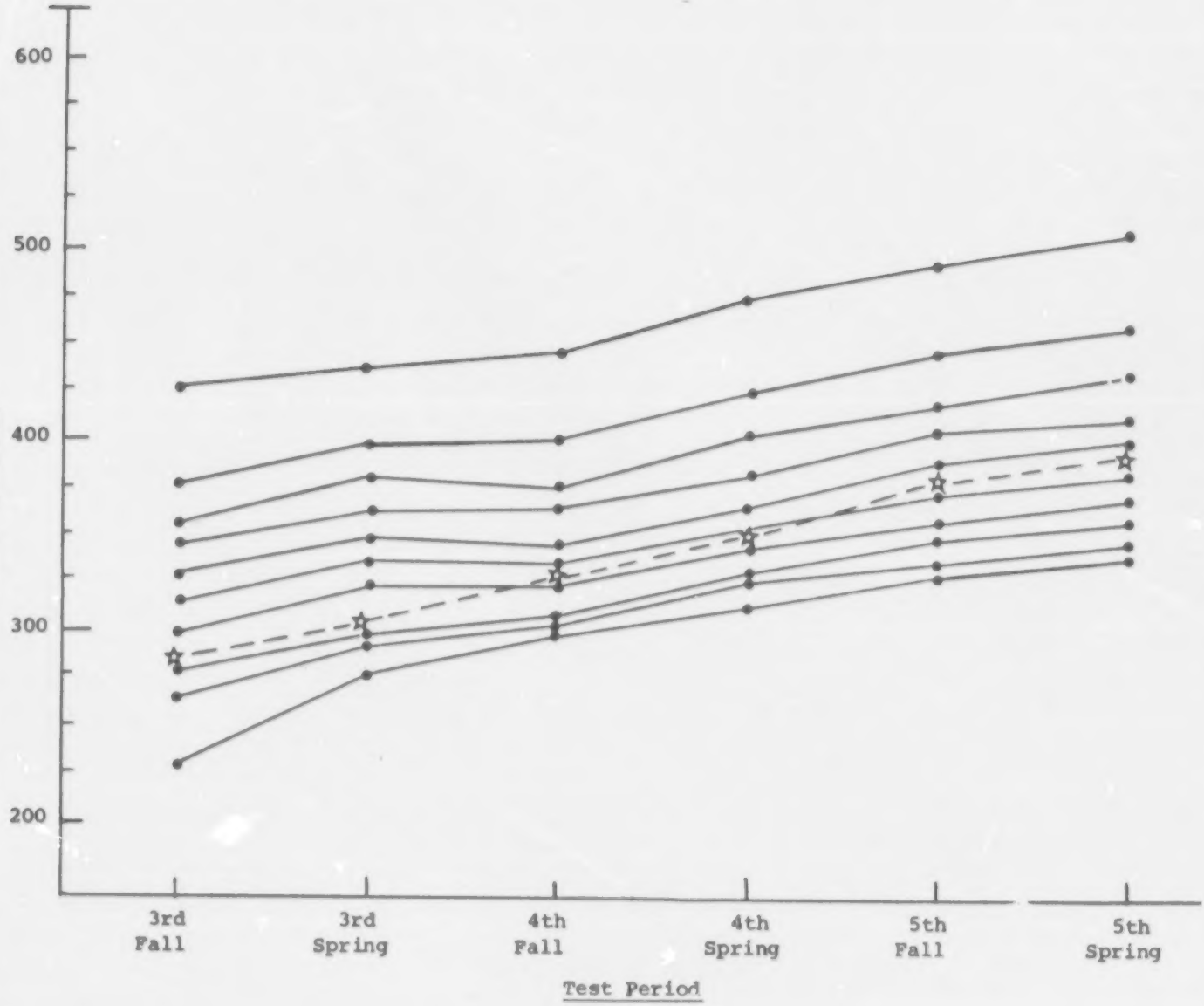


Exhibit 3. The Use of Figures 13-16 for Longitudinal Comparisons

(4) A COMPREHENSIVE LIST OF TABLES

Table 11. Fall, Total Reading Scores: Grade 3, Test Level-2

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1	44	60	1	1	48-49	60
2	2-3	45-46	61-62	2	2-3	50	61
3	4-5	47	63	3	3-5	51	62
4	6-8	48	64	4	6-7	52	63
5	9	49	65	5	9	53	64
6	11	50	66	6	10	54	65
7	13	51	67	7	12	55	66
8	15	52	68	8	14	56	67
9	17	53	69	9	17	57	68
10	19	54	70	10	19	58	69
11	20	55	71	11	20	59	70
12	21	56	72	12	22	60	71
13	23	57	73	13-14	26	61	72
14	25	58-59	74	15	27	62-63	73
15	26	60-61	75	16	30	64	75
16	27	62	76	17	31	65-66	76
17	29	63-64	77	18	32	67	78
18	30	65	78	19	33	68	79
19	31	66-67	79	20	34	69-70	80
20	33	68-69	80	21-22	38	71	81
21-22	34	70	81	23-24	39	72	82
23	35	71	82	25	40	73	83
24	36	72-73	83	26	41	74	84
25	37	74-76	84	27	42	75	85
26	39	77	85	28	43	76	86
27	40	78-79	86	29	44	77	87
28	41	80-82	87	30	45	78-79	88
29	42	83	88	31	47	80-81	89
30	43	84-85	89	32	48	82	90
31	44	86	90	33-34	49	83-84	91
32	45	87-88	91	35-36	50	85	92
33	47	89	92	37	51	86	93
34-35	48	90-91	93	38	52	87-88	94
36	50	92	94	39-40	53	89-90	95
37	51	93-94	95	41	54	91-92	96
38	52	95	96	42	55	93-94	97
39-40	53-54	96	97	43-44	56	95-96	98
41-42	56	97	98	45-46	57	97-99	99
43	57-58	98-99	99	47	59		

89

Table 12. Fall, Total Reading Scores: Grade 4, Test Level-2

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-3	40-41	61	1	1-3	42	61
2	3-5	42	62	2	4-6	43	62
3	6-7	43	63	3	6-7	44	63
4	8-9	44	64	4	8-9	45	65
5	10-11	45-46	66	5-6	10-11	46	66
6	12	47	67	7	13	47	67
7	13	48-49	68	8	15	48-49	69
8	14	50	69	9	17	50	70
9	15-17	51-52	70	10	18	51-52	72
10-11	18	53-54	71	11	20	53	73
12	19	55	72	12	21	54	74
13	20-21	56-57	73	13	23	55	75
14	22	58	74	14	24	56	76
15	24	59	75	15	25	57	77
16	25	60	76	16	26	58	78
17	26	61-62	77	17	27	59	79
18	27	63-64	78	18	29	60	80
19	28-29	65-66	79	19	30	61-62	81
20	31	67	80	20	31	63	82
21	32-33	68-69	81	21	32	64-65	83
22	34	70-71	82	22	34	66	84
23	35-36	72-73	83	23	35	67-68	85
24	38	74-75	84	24	36	69	86
25	39-40	76	85	25	37	70	87
26	42	77	86	26	39	71-72	88
27	44	78	87	27	40	73	89
28	46	79	88	28	42	74-76	90
29	47	80-81	89	29	43	77-79	91
30	48	82-83	90	30-31	44	80-81	92
31	50	84-85	91	32	45	82-83	93
32	52	86	92	33	46	84	94
33	53	87	93	34	48	85-87	95
34	54	88	94	35	50	88-90	96
35	55	89-90	95	36	52	91	97
36	56	91-92	96	37	54	92-93	98
37	58	93-94	97	38-39	55	94-99	99
38	59	95-96	98	40	57		
39	60	97-99	99	41	59		

Table 13. Fall, Total Reading Scores: Grade 4, Test Level-3

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-5	38	65	1	1-3	38	65
2	6-7	39	66	2	4-5	39	66
3	9	40-41	67	3	6	40	67
4	11	42	68	4	8	41	68
5	13	43	69	5	10	42	69
6	16	44-45	70	6	12	43	70
7	20	46	71	7	13	44	71
8	21	47	72	8	16	45	72
9	23	48	73	9	18	46-47	73
10	27	49	74	10	20	48	74
11	28	50	75	11	22	49-50	75
12	30	51-52	76	12	24	51	76
13	32	53-54	77	13	26	52-53	77
14	34	55	78	14	28	54	78
15	35	56	79	15	29	55	79
16	37	57	80	16	31	56	80
17	39	58-59	81	17	33	57-58	81
18	41	60-61	82	18	34	59-61	82
19	42	62-63	83	19	37	62	83
20	44	64	84	20	39	63-64	84
21	46	65	85	21	41	65	85
22	48	66-67	86	22	42	66-67	86
23	49	68-69	87	23	43	68-70	87
24	50	70-71	88	24	44	71-73	88
25	51	72-73	89	25	47	74-75	89
26	52	74	90	26	48	76	90
27	53	75-76	91	27	49	77	91
28	55	77-78	92	28	51	78-79	92
29	56	79-81	93	29	53	80-82	93
30	57	82-83	94	30	54	83-84	94
31	58	84-85	95	31	56	85-86	94-95
32	59	86-88	96	32	58	87-89	96
33	60	89-92	97	33	59	90-91	97
34	61	93-95	98	34	60	92-94	98
35	62	96-99	99	35	62	95-99	99
36	63			36	63		
37	64			37	64		

Table 14. Fall, Total Reading Scores: Grade 5, Test Level-3

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-3	38	65	1	1	38-39	65
2	4-5	39	66	2	2	40	66
3	6	40	67	3	3-5	41	67
4	8-10	41	68	4	6	42	68
5	12	42-43	69	5	8	43	69
6	15	44	70	6	10	44	70
7	17	45	71	7	12	45-46	71
8	20	46	72	8	13	47	72
9	21	47-48	73	9	16	48	73
10	23	49-50	74	10	18	49-50	74
11	25	51	75	11	20	51	75
12	28	52	76	12	22	52	76
13	29	53-54	77	13	24	53-54	77
14	31	55	78	14	29	55	78
15	34	56-57	79	15	31	56-57	79
16	35	58	80	16	33	58	80
17	37	59	81	17	35	59	81
18	38	60-61	82	18	37	60-61	82
19	40	62-63	83	19	39	62-65	83
20	43	64-65	84	20	41	66	84
21	44	66	86	21	43	67-68	85
22	46	67-68	87	22	44	69	86
23	48	69-70	88	23	45	70-71	87
24	49	71-72	89	24	47	72	88
25	50	73-75	90	25	49	73-75	89
26	53	76-78	91	26	51	76	90
27	54	79-80	92	27	52	77-79	91
28	55	81-83	93	28	53	80-81	92
29	56	84-85	94	29	55	82-84	93
30	57	86-88	95	30	56	85	94
31	58	89-91	96	31	57	86-87	95
32	58	92-94	97	32	58	88-90	96
33	60	95-97	98	33	59	91-93	97
34	61	98-99	99	34	60	94-98	98
35	62			35	62	99	99
36	63			36	63		
37	64			37	64		

Table 15. Fall, Total Math Scores: Grade 3, Test Level-2

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-3	39	58	1	1-4	43	56
2	4-5	40-41	59	2	5-6	44	57
3	6-7	42-43	60	3	7-8	45	58
4	8-9	44	61	4	9	46	59
5	10-11	45-46	62	5	11	47	60
6	12	47-48	64	6	12	48	62
7	14	49	65	7	13	49	63
8	15-16	50	67	8	15	50	65
9	18	51	68	9	16	51	66
10	19	52	69	10	17	52	67
11	20	53	70	11	19	53	68
12	21	54-55	71	12	20	54-55	69
13	22-24	56-57	72	13	21	56	70
14	25	58	74	14	23	57	71
15	26	59	75	15	25	58	72
16	28	60	76	16	27	59	73
17	30	61	77	17	28	60	74
18	32	62	78	18	29	61	75
19	33	63	79	19	30	62	76
20	35	64-65	80	20	31	63	77
21	36	66-67	81	21	32	64	78
22	37	68	82	22	33	65	79
23	39	69	83	23	34	66	80
24	42	70	84	24	35	67	81
25	43	71-72	85	25	36	68	82
26	44	73-74	86	26	37	69-70	84
27	45	75	87	27	38	71-73	85
28	46	76	88-89	28	40	74	86
29-30	48	77-78	89	29	41	75	87
31	49	79-81	90	30	42	76-77	89
32	50	82-83	91	31	43	78-79	90
33	51	84	92	32	45	80-82	91
34	52	85-86	93	33	46	83	92
35	53	87-89	94	34	48	84	93
36	54	90-92	95	35	49	85-86	94
37	55	93-95	96	36-37	50	87-89	95
38	57	96-98	97	38-39	51	90-92	96
		99	99	40	52	93-95	97
				41	54	96-97	98
				42	55	98-99	99

Table 16. Fall, Total Math Scores: Grade 4, Test Level-2

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-2	43-44	59	1	1-2	42	59
2	3-4	45	60	2	3-4	43	60
3	5	46	61	3-4	5-6	44-45	61
4	6-7	47	62	5	7	46	62
5	8-9	48	63	6	8-9	47	63
6	9-12	49	64	7	10	48	64
7	13	50	65	8	11-12	49	65
8	14-15	51-52	67	9	14	50-51	67
9	16	53	68	10	15	52-53	68
10	17-18	54	69	11	16	54	69
11	19-20	55	70	12	18	55	70
12	21	56-58	71	13	19	56	71
13	22	59-62	72	14	22	57-58	73
14	24	63	73	15	23	59	74
15	25	64	74	16	24	60	75
16	26	65	75	17	25	61-62	76
17	27-29	66-67	76	18	27	63-64	77
18-19	30	68-69	77	19	29	65-67	78
20	31	70-71	78	20	31	68-70	79
21	32-34	72	79	21-22	32	71	80
22	36-37	73	80	23	34	72-73	81
23	38	74-75	81	24	36	74	82
24-25	39-40	76	82	25	38	75-76	83
26	41	77	83	26	39	77	84
27	42	78-79	84	27	40	78	85
28	43	80	85	28	42	79-80	86
29	44	81	86	29	43	81-82	87
30	46	82-83	87	30	44	83	88
31	47	84	88	31	45	84	89
32	48	85	89	32	48	85-86	90
33	49	86-87	90	33	50	87-88	91
34	50	88-89	91	34-35	51	89-90	92
35	51	90	92	36	52	91-92	93
36	52	91-92	93	37	53	93-94	94
37	53	93-94	94	38-39	54	95-96	95
38	54	95	95	40	56	97	97
39	55	96	96	41	57	98	98
40	56	97	97			99	99
41	57	98	98				
42	58	99	99				

Table 17. Fall, Total Math Scores: Grade 4, Test Level-3

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-5	42-43	61	1	1-5	43	63
2	6-7	44	62	2	6-8	44	65
3	9	45	63	3	10	45	66
4	11	46	64	4	12	46	67
5	14	47	65	5	14	47	68
6	16	48-49	66	6	15	48	69
7	17	50-51	67	7	18	49	70
8	18	52-53	68	8	22	50-51	71
9	22	54	69	9	23	52-54	72
10	23	55	70	10	25	55	73
11	25	56-57	71	11	26	56	74
12-13	28	58	72	12	27	57-59	75
14	29	59-60	73	13	29	60	76
15	30	61	74	14	32	61-62	77
16	31	62-63	75	15	34	63	78
17	33	64-65	77	16	36	64-65	79
18	35	66-67	78	17	38	66-67	80
19	37	68-69	79	18	39	68-70	81
20	38	70	80	19	40	71-72	82
21	39	71-72	81	20	42	73-74	83
22	41	73-74	82	21	44	75	84
23	43	75	83	22	46	76	85
24	44	76	84	23	47	77	86
25	45	77-79	85	24	48	78-80	87
26-27	46	80-82	86	25	49	81-82	88
28	48	83	87	26	50	83	89
29	49	84	88	27-28	51	84-85	90
30	50	85-86	89	29	53	86	91
31	51	87-88	90	30	54	87-88	92
32	52	89-91	91	31-32	55	89-90	93
33	53	92	92	33-34	56	91-92	94
34	55	93	93	35	57	93	95
35-36	56	94	94	36	58	94-95	96
37	57	95	95	37-38	59	96	97
38	58	96-97	96	39-40	60	97-98	97-98
39-40	59	98	96-97	41	61	99	99
41	60	99	98-99	42	62		

Table 18. Fall, Total Math Scores: Grade 5, Test Level-3

FORM A				FORM B			
PUB. ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-3	43	56	1	1-3	40-41	59
2	4-5	44	57	2	4-5	42-43	60
3	6-8	45	58	3	6-8	44-45	61
4	9	46	59	4	10-12	46	62
5	11	47	60	5	13	47	63
6	13	48	61	6	14	48-49	64
7	14	49-50	62	7	17	50	65
8	16	51	63	8	20	51	66
9	18	52-53	64	9	23	52	67
10	19	54	65	10	25	53	68
11	20	55-56	66	11	26	54	69
12	23	57	67	12	27	55	71
13	24	58	68	13	28	56-57	72
14	25	59-60	69	14	29	58	73
15	28	61	70	15	30	59-61	74
16	29	62-63	71	16	31	62-63	75
17	30	64	72	17	32	64-65	76
18	33	65-66	73	18	34	66	77
19	34	67-68	74	19	35	67	78
20-21	35	69-70	75	20	37	68	79
22	37	71	76	21	39	69-70	80
23	38	72-73	77	22	41	71-72	81
24	39	74-75	78	23	42	73-74	82
25	40	76-78	79	24	43	75-76	83
26	41	79	80	25	44	77	84
27	42	80-81	81	26	45	78	85
28	43	82	82	27	46	79-80	86
29	44	83	83	28-29	47	81-82	87
30	45	84	84	30	49	83	88
31	46	85	85	31	50	84-85	89
32	47	86	86	32	51	86-87	90
33	48	87-88	87	33	52	88-89	91
34	49	89-91	88	34-35	53	90-91	92
35	50	92	89	36	54	92-94	93
36	51	93	90	37	55	95	94
37-38	52	94	91	38	56	96	95
39	53	95-96	92	39	57	97	96
40-41	54	97	93			98	97-98
42	55	98	94			99	98-99
		99	95-99				

Table 19. Spring, Total Reading Scores: Grade 3, Test Level-2

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-2	40	58	1	1	38	53
2	2-4	41	59	2	2-3	39	54
3	5-6	42	60	3	4-7	40	55
4	7-8	43	61	4	8	41	56
5	9-10	44	62	5	9-11	42	58
6	11	45	63	6	12	43	60
7	13-14	46	64	7	14	44-45	61
8	15	47-48	65	8	16	46	63
9	16	49	66	9	18	47	65
10	18	50-51	67	10	20	48	66
11	19	52	68	11	22	49	67
12	21	53-54	70	12	24	50	69
13	22	55	71	13	26	51	70
14	23	56-57	72	14	27	52	71
15	25	58-59	73	15	28	53	72
16	26	60-61	74	16	29	54	73
17	27	62-63	75	17	31	55	74
18	28	64	76	18	32	56	76
19	30	65-66	77	19	33	57-58	77
20	31	67	78	20	34	59-60	78
21	32	68	79	21	35	61	79
22	33	69-70	80	22	36	62-63	80
23	34	71	81	23	37	64	81
24	35	72	82	24	38	65	82
25	37	73-74	83	25	39	66	83
26	38-39	75-76	84	26	40	67-68	84
27	41	77	85	27	41	69-70	85
28	42	78-79	86	28	42	71	86
29	43	80	87	29	43	72-73	87
30-31	44-46	81	88	30	44	74	88
32	47-49	82	89	31	45	75-76	89
33-34	51	83	90	32	46	77	90
35	52	84	91	33	47	78-79	91
36	53	85-86	92	34	48	80-81	92
37	55	87-88	93	35	49	82-83	93
38	56	89-90	94	36	50	84-85	94
39	57	91-92	95	37	51	86-88	95
		93-94	96			89-90	96
		95	97			91-93	97
		96-97	98			94-96	98
		98-99	99			97-99	99

Table 20. Spring, Total Reading Scores: Grade 4, Test Level-3

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-2	38	64	1	1	38	66
2	3	39	65	2	2-3	39-40	67
3	4-6	40	66	3	4	41	68
4	7	41	67	4	5	42	69
5	10	42	68	5	7	43	70
6	12	43	69	6	9	44-45	71
7	15	44	70	7	12	46	72
8	17	45-46	71	8	14	47-48	73
9	18	47	72	9	15	49	74
10	21	48	73	10	17	50	75
11	23	49-50	74	11	18	51	76
12	25	51	75	12	20	52-53	77
13	29	52-53	76	13	22	54-55	78
14	31	54	77	14	24	56	79
15	33	55	78	15	26	57	80
16	35	56	79	16	29	58-59	81
17	36	57	80	17	31	60-61	82
18	40	58	81	18	33	62-63	83
19	41	59-60	82	19	35	64-65	84
20	43	61-62	83	20	37	66	85
21	45	63	84	21	39	67-68	86
22	46	64	85	22	40	69-70	87
23	47	65-66	86	23	43	71-73	88
24	48	67-69	87	24	45	74	89
25	49	70-71	88	25	47	75-76	90
26	51	72-73	89	26	50	77-78	91
27	52	74-75	90	27	52	79-81	92
28	53	76-78	91	28	53	82-83	93
29	54	79-80	92	29	54	84-85	94
30	56	81	93	30	56	86-87	95
31	57	82-85	94	31	57	88-90	96
32	58	86	95	32	58	91-93	97
33	59	87-89	96	33	60	94-96	98
34	60	90-92	97	34	62	97-99	99
36	61	93-97	98	35	63		
37	63	98-99	99	37	65		

Table 21. Spring, Total Reading Scores: Grade 5, Test Level-3

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-2	38	67	1	1	39	67
2	2-3	39	68	2	2	40	68
3	6-7	40	69	3	3-5	41	69
4	9	41-42	70	4	7	42	70
5	11-13	43	71	5	9	43-44	71
6	16	44-45	73	6	11	45	73
7	18	46	74	7	14	46-47	74
8	20	47	75	8	18	48-49	75
9	23	48-49	76	9	21	50	76
10	25	50	77	10	23	51-52	77
11	28	51	78	11	25	53-54	78
12	31	52-53	79	12	29	55-56	79
13	33	54	80	13	33	57	80
14	35	55	81	14	35	58	81
15	36	56	82	15	36	59	82
16	38	57-58	83	16	38	60	83
17	39	59-60	84	17	40	61	84
18	40	61-62	85	18	42	62-63	85
19	43	63-64	86	19	43	64	86
20	45	65-66	87	20	46	65-66	87
21	46	67-68	88	21	48	67	88
22	47	69-70	89	22	50	68-70	89
23	48	71-72	90	23	52	71-72	90
24	49	73-74	91	24	53	73-74	91
25	52	75-77	92	25	54	75-77	92
26	53	78-80	93	26	55	78-80	93
27	54	81-83	94	27	56	81-83	94
28	56	84-86	95	28	57	84-86	95
29	57	87-90	96	29	58	87-89	96
30	58	91-93	97	30	59	90-92	97
31	59	94-95	98	31	60	93-94	98
32	60	96-99	99	32	61	95-99	99
33	61			33	62		
34	62			34-35	63		
35	63			36	64		
36	64			37	65		
37	66			38	66		

Table 22. Spring, Total Math Scores: Grade 3, Test Level-2

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-3	44	56	1	1-3	46	58
2	3-5	45	57	2	4-5	47	60
3	5-7	46-47	58	3	5-7	48	62
4	8-9	48-49	60	4	8	49	63
5	9-10	50	61	5	9-10	50	64
6	10-12	51-52	62	6	12	51-52	65
7	13	53	63	7-8	13-15	53-54	66
8	14	54	64	9	16	55-56	67
9	15-16	55	65	10	17	57-58	69
10	17	56-57	66	11	19	59	70
11	18-19	58	67	12	21	60-61	72
12	20	59	68	13	23	62-64	73
13-14	21	60	69	14	24	65	74
15	22-23	61	70	15-16	26	66	75
16	25	62	71	17	27	67	76
17	26	63	72	18	28	68-69	77
18	27	64-65	73	19	29	70	78
19	28	66-67	74	20	30	71	79
20	29	68-70	75	21	31	72	80
21	30	71-72	76	22	32	73-74	81
22	31	73-74	77	23	33	75-76	82
23	32	75-76	78	24	35	77	83
24	33	77	79	25	38	78-79	84
25	34	78	80	26	39	80-81	85
26	35	79-80	81	27	40	82-83	86
27-28	36	81-82	82	28	41	84-85	87
29	38	83	83	29-30	42	86-88	88
30	39	84	84	31	43	89-90	89
31	40	85	85	32	45	91	90
32-33	42	86	86	33	46	92-93	91
34	43	87-88	87	34-35	47	94	92
35	45	89	88	36	48	95	93
36	46	90	89	37	50	96	95
37	47	91	90	38	52	97	96
38	48	92-93	91	39-41	54	98	97
39	49	94	92	42-43	55	99	98-99
40	51	95	93	44-45	57		
41	52	96	94				
42	53	97	95				
43	55	98	96				
		99	97-99				

Table 23. Spring, Total Math Scores: Grade 4, Test Level-3

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-3	42-43	56	1	1-3	42-43	60
2	4-6	44	57	2	4-6	44	61
3	7-8	45	58	3	8	45-46	62
4	10	46	59	4	10	47	63
5	12	47	60	5	12	48-49	64
6	15	48	61	6	14	50	65
7	16	49-50	62	7	17	51	67
8	17	51-52	63	8	20	52-53	68
9	19	53-54	64	9	22	54	69
10	21	55-56	65	10	24	55	70
11	22	57-58	66	11	28	56-57	71
12	24	59	67	12	29	58	72
13	27	60	68	13	30	59	73
14	29	61	69	14	32	60-61	74
15	30	62	70	15	33	62-63	75
16	31	63	71	16	34	64-65	76
17	32	64	72	17	35	66	77
18	34	65-66	73	18-19	37	67-68	78
19	35	67-68	74	20	38	69-70	79
20	37	69-70	75	21	39	71	80
21	38	71-72	76	22	40	72	81
22	39	73-75	77	23	42	73-74	82
23	40	76-77	78	24	43	75	83
24	41	78	79	25	44	76-77	84
25	42	79-80	80	26	46	78-79	85
26-27	43	81	81	27	47	80-81	86
28	45	82-83	82	28	48	82-83	87
29	46	84	83	29	49	84-85	88
30	47	85	84	30-31	50	86-87	89
31-32	48	86-87	85	32	51	88-89	90
33	49	88-89	86	33	52	90	91
34	50	90-91	87	34-35	53	91	92
35	51	92	88	36	54	92-93	93
36	52	93-94	89	37	55	94	94
37-38	53	95	91	38-39	56	95	95
39	54	96	92	40	57	96-97	96
40-41	55	97	93	41	59	98	97
		98	94			99	97-99
		99	95-99				

Table 24. Spring, Total Math Scores: Grade 5, Test Level-3

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-3	43	58	1	1-3	41-42	63
2	4-6	44	59	2	5-6	43	64
3	7-8	45-46	60	3	8-9	44-45	65
4	10	47	61	4	11	46	66
5	11-13	48	62	5	13-16	47	67
6	14	49	63	6	18	48	68
7	16	50	64	7	20	49-50	69
8	18	51	65	8	23	51	70
9	19	52-53	66	9	24	52-53	71
10	21	54-55	67	10	25	54	72
11	23	56-57	68	11	26	55-56	73
12	25	58-59	69	12	28	57	74
13	27	60-61	70	13	29	58	75
14	28	62	71	14	30	59-60	76
15	29	63-64	72	15	33	61	77
16-17	31	65	73	16	36	62-63	78
18	32	66-67	74	17	37	64-65	79
19	33	68-69	75	18-19	38	66	80
20	36	70-72	76	20	39	67-68	81
21	37	73-74	77	21	40	69	82
22	38	75	78	22	41	70-71	83
23-24	40	76	79	23	42	72-73	84
25	41	77-78	80	24	44	74-75	85
26	42	79	81	25	46	76-77	86
27-28	45	80-81	82	26	47	78-79	87
29	46	82-83	83	27	49	80-81	88
30-31	47	84	84	28	50	82-83	89
32	48	85-86	85	29	51	84-85	90
33	49	87-88	86	30	52	86-87	91
34	50	89	88	31	53	88-89	92
35	51	90-91	89	32	55	90	93
36	52	92	90	33	57	91	94
37	53	93	91	34-35	58	92	95
38-39	54	94	92	36	59	93-95	96
40	55	95	93	37-38	60	96-97	97
41	56	96	94	39	61	98-99	98
42	57	97	95	40	62		99
		98	97				
		99	98-99				

Table 25. Fall, Total Reading Scores: Grade 10, Test Level-2

FORM E				FORM F			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-3	42	69	1	1-3	41	69
2	4-6	43	70	2	4-6	42	70
3	7	44	71	3	8	43-44	71
4	8	45	72	4	10-12	45	72
5	10-12	46	73	5	14	46	73
6	15	47-48	74	6	16-18	47-48	74
7	18	49-50	75	7	21	49-50	75
8	19	51-52	76	8	24	51-52	76
9	20	53	77	9	26	53	77
10	23	54	78	10	29	54	78
11	26	55	80	11	32	55-56	79
12	29	56-58	81	12	33	57-58	80
13	32	59	82	13	34	59	81
14	34	60-61	83	14	36	60	83
15	35	62-63	84	15	37	61-62	84
16	38	64	85	16	38-41	63-64	85
17	40	65-66	86	17	42	65	86
18	43	67-69	87	18	44	66	87
19	44	70-71	88	19	45	67-69	88
20	45	72-73	89	20	46	70-71	89
21	47	74	90	21	48	72-74	90
22	48	75-76	91	22	49	75-76	91
23	49	77-79	92	23	50	77-79	92
24	51	80	93	24	52	80-81	93
25	53	81-83	94	25	53	82-84	94
26	54	84-85	95	26	54	85-86	95
27	56	86-87	96	27	56	87-89	96
28	58	88-89	97	28	57	90-91	97
29	59	90-93	98	29	58	92-94	98
30	60	94-99	99	30	60	95-99	99
31-32	61			31-32	61		
33-34	63			33	62		
35-36	64			34	63		
37	65			35-36	64		
38	66			37	65		
39-40	67			38	67		
41	68			39-40	68		

Table 26. Fall, Total Reading Scores: Grade 11, Test Level-2

FORM E				FORM F			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-3	43	67	1	1-3	39	64
2	4-5	44	68	2	4-5	40	65
3	6-7	45	69	3	6-7	41	66
4	9-10	46	70	4	9	42-43	67
5	12	47	72	5	11-13	44	68
6	16	48	73	6	15-17	45	69
7	17	49-50	74	7	19	46	70
8	19	51-52	75	8	21	47	71
9	21	53	76	9	22	48	72
10	23	54	77	10	24	49	73
11	26	55	78	11	26	50-51	74
12	27	56-57	79	12	29	52-53	75
13	29	58-60	80	13	31	54	76
14	31	61-62	81	14-15	33	55	77
15	32	63	82	16	34-36	56-57	79
16	33-35	64-66	83	17	37	58-59	80
17	36	67	84	18	38	60-61	81
18	37	68-69	85	19	39	62	82
19	39	70-71	86	20	41	63-64	83
20	41	72-74	87	21	43	65-66	84
21	43	75	88	22	44	67	85
22-23	45	76	89	23	45	68-70	86
24	46	77-78	90	24	46	71-73	87
25	48	79-80	91	25	48	74	88
26	50	81-82	92	26	50	75-76	89
27-28	51	83-85	93	27	51	77-78	90
29	53	86-87	94	28	52	79-80	91
30-31	55	88-89	95	29	53	81-83	92
32	56	90-91	96	30	54	84-85	93
33	58	91-94	97	31	55	86-87	94
34	59	95-96	98	32	57	88-89	95
35	60	97-99	99	33	58	90-91	96
36-37	61			34	59	92-93	97
38	62			35	60	94	98
39	64			36	61	95-99	99
40-41	65			37	62		
42	66			38	63		

Table 27. Fall, Total Reading Scores: Grade 12, Test Level-2

FORM E				FORM F			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-4	40	64	1	1-3	41	62
2	4-5	41-42	66	2	4-6	42-43	63
3	7-8	43-44	67	3	7-8	44	65
4	9-11	45	68	4	10-11	45-46	66
5	12	46	69	5	12-14	47	68
6	14-16	47-48	70	6	16	48-49	69
7	18	49	71	7	17-18	50	70
8	19	50-51	72	8	20	51	72
9	21-23	52-53	73	9	21	52-53	73
10	26	54-55	74	10	22	54-55	74
11	28	56	75	11	23	56	75
12	29	57	76	12	25	57	76
13	30	58	77	13	27	58-59	77
14	32	59-60	78	14	28	60-61	78
15	33	61-62	79	15	30	62-63	79
16	35	63-64	80	16	32	64-65	80
17	36-37	65	81	17	33	66	81
18	38	66-67	83	18	35	67	82
19	39	68-69	84	19	36	68	83
20	41	70-71	85	20	37	69-70	84
21	42	72	86	21	39	71-72	85
22	44	73-74	87	22	40	73	86
23	45	75-76	88	23	42	74-75	87
24-25	47	77-78	89	24	44	76	88
26	48	79-80	90	25	46	77	89
27	50	81-82	91	26	47	78-79	90
28	51	83-84	92	27	48	80-81	91
29	52	85-86	93	28	49	82-83	92
30	54	87-88	94	29	50	84-85	93
31	55	89	95	30	51	86	94
32-33	56	90	96	31	53	87-88	95
34	57	91-92	97	32-33	54	89-91	96
35	58	93-94	98	34	55	92-94	97
36	59	95-99	99	35-36	56	95-96	98
37	60			37	58	97-98	99
38	62			38	59		
39	63			39-40	60		

Table 28. Fall, Total Math Scores: Grade 10, Test Level-5

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-4	41	76	1	1-5	43	75
2	5-8	42-43	77	2	6-11	44-45	76
3	10-13	44-45	78	3	13-15	46-47	77
4	15-18	46-47	79	4	17-19	48-49	78
5	20	48-49	80	5	21-23	50-51	79
6	23-25	50-51	81	6	25	52	80
7	27	52	82	7	27-30	53-54	81
8	29-32	53-55	83	8	33	55-56	82
9	34	56-57	84	9	34	57-58	83
10	37	58-59	85	10	36	59-60	84
11	40	60-61	86	11	39	61-62	85
12	42	62-64	87	12	41	63	86
13	44	65-66	88	13	43-45	64	87
14	47	67-69	89	14	46	65-66	88
15	49	70-73	90	15	47	67-69	89
16	50	74-75	91	16	48	70-71	90
17	52	76-77	92	17	50	72-73	91
18	53	78-80	93	18	52	74-75	92
19	54	81-84	94	19	54	76-78	93
20	55	85-88	95	20	55	79-81	94
21	57	89-91	96	21	56	82-83	95
22	58	92-94	97	22	57	84-86	96
23	59	95-97	98	23	59	87-90	97
24	60	98-99	99	24	60	91-95	98
25	61			25	61	96-99	99
26	62			26	62		
27	63			27	63		
28	65			28-29	64		
29-30	67			30	66		
31-32	68			31-32	67		
33	69			33-34	68		
34-35	70			35	69		
36	71			36	70		
37	72			37	71		
38	73			38	72		
39	74			39-40	73		
40	75			41-42	74		

Table 29. Fall, Total Math Scores: Grade 11, Test Level-5

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-8	45-46	76	1	1-8	47-48	77
2	9-12	47-48	77	2	9-14	49-50	78
3	13-18	49	78	3	16-18	51-52	79
4	20-22	50-51	79	4	19-21	53	80
5	23-25	52-54	80	5	23-24	54	81
6	27-30	55-56	81	6	26-28	55-56	82
7	32	57-58	82	7	30	57-58	83
8	33	59-61	83	8	32-34	59-60	84
9	35-37	62-65	84	9	35	61-62	85
10	39	66-67	85	10	37	63-64	86
11	41	68-69	86	11	39	65	87
12	43	70-71	87	12	40-42	66-67	88
13	45	72-73	88	13	43	68-69	89
14	47	74-75	89	14	45	70-72	90
15	48	76-77	90	15	46	73-75	91
16	50	78-79	91	16	48	76-79	92
17	52	80-83	92	17-18	49	80-81	93
18	54	84-85	93	19	51	82	94
19	55	86-87	94	20	52	83-85	95
20	56	88-91	95	21	54	86-89	96
21	57	92-93	96	22	56	90-93	97
22	58	94-95	97	23-24	57	94	98
23-24	60	96-98	98	25	59	96-99	99
25	61	99	99	26-27	60		
26	62			28	62		
27-28	64			29-30	63		
29	65			31	64		
30-31	66			32	65		
32	67			33	66		
33-34	68			34	67		
35-36	69			35-36	68		
37	70			37	70		
38-39	71			38-39	71		
40	72			40-41	72		
41	73			42-43	73		
42	74			44	75		
43-44	75			45-46	76		

Table 30. Fall, Total Math Scores: Grade 12, Test Level-5

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-13	44-45	77	1	1-8	48-50	76
2	15-19	46-47	78	2	10-15	51-53	77
3	20-22	48-49	79	3	17-19	54	78
4	23-25	50-51	80	4	20-23	55-56	79
5	27-28	52	81	5	25-26	57-58	80
6	30-32	53-54	82	6	28	59-60	81
7	34-36	55-57	83	7	30-33	61-62	82
8	38	58-59	84	8	35	63-64	83
9	39	60-61	85	9	36-39	65	85
10	41-43	62-63	86	10	40	66-67	86
11	44	64-65	87	11	41	68-69	87
12	46	66-67	88	12	43	70	88
13	47	68	89	13	44	71-72	89
14	48	70-73	90	14	46	73-75	90
15	50	74-76	91	15	48	76-77	91
16	52	77-78	92	16	50	78-79	92
17-18	53	79-80	93	17	51	80-82	93
19	54	81-83	94	18	53	83-85	94
20	56	84-87	95	19	54	86-87	95
21	57	88-90	96	20	55	88-89	96
22	58	91-93	97	21	57	90-92	97
23	59	94-96	98	22-23	58	93-95	98
24-25	60	97-99	99	24	60	96-99	99
26	62			25	61		
27	63			26-27	62		
28	64			28	63		
29	65			29-30	65		
30-31	66			31-32	66		
32	67			33-34	67		
33	68			35-36	68		
34	69			37	69		
35-36	70			38-39	70		
37	71			40	71		
38-39	73			41-42	72		
40	74			43-44	73		
41-42	75			45-46	74		
43	76			47	75		

Table 31. Spring, Total Reading Scores: Grade 10, Test Level-2

FORM E				FORM F			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-5	40-41	66	1	1-3	40-41	66
2	5-7	42	68	2	4-5	42-43	67
3	8	43-44	70	3	6	44-45	69
4	10	45	71	4	8-10	46	70
5	12-14	46	72	5	12	47-48	71
6	17	47-48	73	6	15-18	49-50	72
7	19	49-50	74	7	20	51	73
8	21	51-52	75	8	23	52-53	74
9	22	53	76	9	25	54	75
10	24	54	77	10	28	55-56	76
11	27	55	78	11	30	57-58	77
12	29	56-57	80	12	31	59	78
13	32	58	81	13	33	60	79
14	34	59	82	14	35	61-62	80
15	35	60-61	83	15	36	63-64	81
16	37	62-63	84	16	37-40	65-66	82
17	40	64	85	17	41	67-68	83
18	42	65-66	86	18	42	69-70	84
19	43	67-68	87	19	43	71-72	85
20	44	69-71	88	20	44	73	86
21	46	72	89	21	46	74	87
22	47	73	90	22-23	47	75-76	88
23	48	74-75	91	24	49	77-78	89
24	50	76-79	92	25	50	79-80	90
25	51	80	93	26	51	81-83	91
26	52	81-83	94	27	52	84	92
27	53	84-85	95	28	53	85-86	93
28	55	86-89	96	29	54	87-89	94
29	56	90-91	97	30	55	90	95
30-31	57	92	98	31	56	91-92	96
32	58	93-99	99	32	57	93	97
33	60			33	59	94	98
34	61			34	61	95-99	99
35	62			35	62		
36	63			36-37	63		
37	64			38	64		
38-39	65			39	65		

Table 32. Spring, Total Reading Scores: Grade 11, Test Level-2

FORM E				FORM F			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-3	39-40	64	1	1-3	45	68
2	4-5	41	65	2	4-5	46-47	69
3	6-8	42-43	66	3	7-9	48	70
4	10-12	44	67	4	10	49-51	72
5	14	45	68	5	12-14	52-53	73
6	16	46	69	6	16-18	54	74
7	18	47	70	7	20	55	75
8	21	48	72	8	22	56	76
9	23	49-50	73	9	25	57-58	77
10	26	51-52	74	10	27	59-60	78
11	28	53	75	11	29	61	79
12	30	54	76	12	31	62	80
13	32	55	77	13	33	63	81
14	34	56-57	78	14-15	35	64-65	82
15	35	58-59	79	16	36-38	66-67	83
16	35-37	60	80	17	38	68-69	84
17	38	61	81	18	39	70-72	85
18	39	62	82	19	41	73-75	86
19	41	63-65	83	20	42	76-77	87
20	42	66	84	21	44	78	88
21	44	67-69	85	22	45	79-80	89
22	46	70-71	86	23-24	46	81-83	90
23	47	72-73	87	25	48	84-85	91
24	48	74-76	88	26	49	86	92
25	49	77	89	27	51	87-88	93
26	50	78-79	90	28-29	52	89-91	94
27	51	80-82	91	30	54	92	95
28	52	83-84	92	31	55	93-94	96
29	53	85	93	32	56	95	97
30	54	86-87	94	33-34	57	96-97	98
31	55	88-89	95	35	59	98-99	99
32	56	90-91	96	36-37	60		
33	57	92-93	97	38	62		
34	58	94	98	39-40	63		
35-36	60	95-99	99	41	64		
37	61			42-43	66		
38	62			44	67		

Table 33. Spring, Total Reading Scores: Grade 12, Test Level-2

FORM E				FORM F			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-5	40	67	1	1-4	42-43	67
2	6-7	41	68	2	5-7	44	68
3	8-10	42	69	3	9-11	45-46	69
4	12-13	43-44	70	4	13-14	47	70
5	14	45	71	5	16-18	48	71
6	16-18	46	72	6	20	49-50	72
7	20	47-49	73	7	22-24	51	73
8	22	50-52	74	8	26	52-55	74
9	23-25	53-54	75	9	27	56	76
10	27	55	76	10	29	57	77
11	29	56	77	11	30	58-59	78
12	30	57-58	78	12	32	60-61	79
13	32	59	79	13	34	62	80
14	33	60	80	14	36	63	81
15	35	61-62	81	15	38	64-65	82
16	37	63-64	82	16	39	64-68	83
17	39-41	65	83	17	41	69-70	84
18	41	66-67	84	18	43	71	85
19	42	68-69	85	19	45	72	86
20	44	70-71	86	20	46	73-75	87
21	46	72	87	21	48	76	88
22	48	73-74	88	22	50	77-78	89
23	50	75-76	89	23	51	79-81	90
24	51	77-78	90	24	52	82	91
25	52	79-80	91	25-26	53	83	92
26	53	81-82	92	27	54	84	93
27	54	83-84	93	28	55	85-86	94
28	55	85-86	94	29	56	87-88	95
29	57	87-89	95	30	57	89-90	96
30	58	90-91	96	31	59	91-93	97
31	59	92-93	97	32-33	60	94-96	98
32-33	60	94-96	98	34-35	61	97-99	99
34-35	61	97-99	99	36	62		
36	62			37	63		
37	63			38	64		
38	65			39-40	65		
39	66			41	66		

Table 34. Spring, Total Math Scores: Grade 10, Test Level-5

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-8	49	80	1	1-7	46-47	77
2	9-15	50-51	81	2	9-12	48-49	78
3	17-19	52	82	3	14-16	50	79
4	21-23	53-54	83	4	18-20	51	80
5	25-27	55-56	84	5	22-23	52-53	81
6	30-32	57-58	85	6	26-28	54-57	82
7	34	59	86	7	30	58-59	83
8	36	60	87	8	32-34	60-61	84
9	38-40	61-63	88	9	36	62-63	85
10	42	64-65	89	10	38	64-66	86
11	43	66-69	90	11	40-41	67-68	87
12	45	70-71	91	12	43	69-70	88
13	47	72-75	92	13	45	71-74	89
14	49	76-77	93	14	47	75	90
15	51	78-80	94	15	49	76-77	91
16	53	81-84	95	16	50	78-81	92
17	55	85-88	96	17	51	82-83	93
18	56	89-92	97	18	52	84-85	94
19-20	57	93-95	98	19	54	86-88	95
21	58	96-99	99	20	56	89-91	96
22	60			21	58	92-94	97
23-24	62			22	59	95-97	98
25	63			23-24	60	98-99	99
26-27	64			25-26	61		
28	65			27	63		
29-30	66			28	64		
31	67			29	65		
32-33	69			30	66		
34	70			31	67		
35	71			32-33	68		
36	72			34-35	69		
37-39	73			36	71		
40-41	74			37	72		
42-43	76			38-40	73		
44-45	77			41-42	74		
46-47	78			43-44	75		
48	79			45	76		

Table 35. Spring, Total Math Scores: Grade 11, Test Level-5

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-10	46-47	77	1	1-8	44	74
2	12-15	48-49	78	2	10-15	45-46	75
3	16-20	50	79	3	17-18	47	76
4	22-24	51-54	80	4	20-24	48-50	77
5	26-28	55-56	81	5	26-28	51	78
6	29-31	57	82	6	30	52-53	79
7	34	58-59	83	7	31-33	54	80
8	36-37	60-61	84	8	35	55-58	81
9	39	62-63	85	9	36-39	59-60	82
10	41	64-65	86	10	40	61-62	83
11	42-43	66-67	87	11	42	63-64	84
12	45	68-69	88	12	44	65-66	85
13	47	70-71	89	13	45	67	86
14	48	72-73	90	14	46	68-69	87
15	50	74-77	91	15	48	70-71	88
16-17	51	78-80	92	16	50	72-75	89
18	52	81-82	93	17	51	76-78	90
19	53	83-84	94	18	52	79-81	91
20	55	85-88	95	19	54	82-83	92
21	57	89-90	96	20	55	84-86	93
22	58	91-94	97	21	56	87-88	94
23	59	95-96	98	22	57	89-90	95
24	60	97-99	99	23	58	91-92	96
25	61			24	59	93-95	97
26-27	63			25	60	96-97	98
28	64			26-27	61	98-99	99
29	65			28	62		
30-32	67			29-30	63		
33-34	68			31	64		
35-36	69			32-33	66		
37	70			34	67		
38-39	71			35-36	68		
40	72			37-38	69		
41	73			39	70		
42	74			40-41	71		
43-44	75			42	72		
45	76			43	73		

Table 36. Spring, Total Math Scores: Grade 12, Test Level-5

FORM A				FORM B			
PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE	PUB. %ILE	ESAA %ILE
1	1-16	50-51	80	1	1-14	46-47	77
2	17-21	52-53	81	2	15-18	48	78
3	22-25	54-55	82	3	20-23	49-50	79
4	27-29	56-57	83	4	25-27	51-52	80
5	31-32	58-59	84	5	30-32	53-54	81
6	34-36	60-61	85	6	33-35	55	82
7	38	62-64	86	7	37-38	56-57	83
8	40-42	65-67	87	8	39	58	84
9	44	68-69	88	9	41	59-61	85
10	46	70-73	89	10	43-44	62-64	86
11	47	74-76	90	11	46	65-67	87
12	49	77-78	91	12	48	68-69	88
13	50	79-81	92	13	49	70-72	89
14	52	82-83	93	14	50	73-75	90
15	53-55	84-86	94	15	52	76-78	91
16	56	87-89	95	16	53	79-81	92
17	57	90-91	96	17	54	82-84	93
18	58	92-93	97	18	56	85-86	94
19	60	94-96	98	19	57	87-88	95
20	62	97-99	99	20-21	58	89-90	96
21-22	63			22	59	91-92	97
23	64			23	60	93-94	98
24	65			24	61	95-99	99
25	66			25-26	62		
26-27	67			27	64		
28	68			28-29	65		
29-30	69			30	66		
31	70			31-32	67		
32-33	71			33	68		
34	72			34-35	69		
35-37	73			36	70		
38-39	74			37-38	71		
40-41	75			39-40	72		
42-43	76			41	73		
44	77			42-43	74		
45-47	78			44	75		
48-49	79			45	76		

Table 37. Total Reading Scores: Pilot, Grade 3, Test Level-2

PUB. Percentile	Form A		Form B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	12	11	11	12
10	26	23	26	27
15	34	33	36	37
20	42	41	45	45
25	48	48	50	50
30	54	58	56	56
35	60	65	61	62
40	65	71	66	68
45	73	76	69	76
50	78	80	73	81
55	82	82	78	85
60	86	86	82	89
65	88	88	86	92
70	90	89	90	94
75	92	91	92	96
80	93	93	94	97
85	95	96	97	98
90	96	98	99	99
95	99	99	99	99

Table 38. Total Reading Scores: Pilot, Grade 4, Test Level-2

PUB. Percentile	Form A		Form B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	14	Level 2 not used in spring	14	Level 2 not used in spring
10	24		16	
15	32		33	
20	42		40	
25	52		48	
30	62		55	
35	67		64	
40	72		69	
45	78		76	
50	80		80	
55	82		85	
60	87		89	
65	89		91	
70	90		94	
75	92		96	
80	94		96	
85	96		98	
90	98		99	
95	99		99	

Table 39. Total Reading Scores: Pilot, Grade 4, Test Level-3

PUB. Percentile	Form A		Form B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	15	11	12	9
10	31	26	24	18
15	41	40	34	31
20	51	52	47	42
25	63	60	53	54
30	66	65	65	64
35	72	72	75	72
40	77	76	78	77
45	81	81	83	82
50	85	84	85	84
55	86	87	88	88
60	90	92	91	90
65	92	93	93	92
70	94	95	94	93
75	96	96	96	95
80	97	97	98	96
85	99	98	98	98
90	99	99	99	99
95	99	99	99	99

Table 40. Total Reading Scores: Pilot, Grade 5, Test Level-3

PUB. Percentile	Form A		Form B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	16	15	11	11
10	28	33	21	26
15	41	45	36	44
20	52	56	51	56
25	58	63	60	64
30	66	69	66	72
35	73	73	72	75
40	77	79	76	79
45	82	83	82	83
50	86	87	85	86
55	88	89	87	88
60	90	92	90	91
65	92	94	92	94
70	95	96	94	96
75	97	96	95	97
80	97	97	97	98
85	98	98	98	99
90	99	99	99	99
95	99	99	99	99

Table 41. Total Math Scores: Pilot, Grade 3, Test Level-2

PUB. Percentile	Form A		Form B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	12	10	12	10
10	23	22	20	22
15	32	27	28	31
20	42	34	34	36
25	50	41	40	47
30	56	47	47	52
35	61	56	57	59
40	67	61	60	64
45	68	67	66	68
50	76	71	72	74
55	79	73	78	77
60	84	78	83	80
65	88	82	88	83
70	90	84	91	87
75	93	85	93	88
80	94	87	95	90
85	96	91	97	93
90	97	93	98	95
95	99	96	99	97

Table 42. Total Math Scores: Pilot, Grade 4, Test Level-2

PUB. Percentile	Form A		Form B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	9	Level 2	7	Level 2
10	18	not used	16	not used
15	29	in spring	23	in spring
20	36		32	
25	43		37	
30	51		45	
35	58		53	
40	62		56	
45	65		65	
50	70		72	
55	76		75	
60	76		79	
65	81		81	
70	85		84	
75	88		87	
80	90		91	
85	93		92	
90	95		95	
95	95		97	

Table 43. Total Math Scores: Pilot, Grade 4, Test Level-3

PUB. Percentile	Form A		Form B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	17	13	17	13
10	26	24	30	27
15	33	33	40	35
20	42	39	44	43
25	50	49	55	51
30	57	51	59	58
35	63	56	64	61
40	66	59	71	65
45	69	67	74	71
50	73	69	77	73
55	75	71	80	76
60	80	76	83	81
65	84	80	87	83
70	85	82	89	87
75	88	83	90	91
80	90	86	94	92
85	93	90	96	94
90	95	92	98	96
95	97	95	99	97

Table 44. Total Math Scores: Pilot, Grade 5, Test Level-3

PUB. Percentile	Form A		Form B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	12	14	15	15
10	20	24	27	30
15	31	31	34	39
20	36	41	45	46
25	45	48	52	54
30	51	54	56	60
35	56	59	62	63
40	61	63	68	68
45	64	67	70	73
50	69	70	75	78
55	72	72	78	82
60	77	76	82	84
65	82	78	84	88
70	83	80	88	89
75	85	84	91	92
80	88	88	92	94
85	91	89	94	96
90	94	92	96	98
95	96	96	98	99

Table 45. Total Reading Scores: Basic, Grade 3, Test Level-2

PUB. Percentile	Form A		Form B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	8	8	7	8
10	15	14	15	16
15	21	20	22	22
20	28	25	28	28
25	32	30	32	33
30	36	37	38	37
35	42	45	41	42
40	48	49	46	47
45	54	56	48	54
50	60	60	53	62
55	64	63	60	68
60	68	68	64	73
65	71	71	69	76
70	75	75	75	81
75	79	78	79	85
80	81	82	85	90
85	86	90	90	91
90	88	92	93	95
95	95	97	97	98

Table 46. Total Reading Scores: Basic, Grade 4, Test Level-2

PUB. Percentile	Form A		Form B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	8	Level 2	7	Level 2
10	14	not used	14	not used
15	19	in spring	19	in spring
20	24		24	
25	31		31	
30	40		35	
35	47		42	
40	54		50	
45	59		60	
50	62		64	
55	64		68	
60	70		73	
65	73		78	
70	77		84	
75	81		86	
80	86		88	
85	90		92	
90	94		94	
95	96		99	

Table 47. Total Reading Scores: Basic, Grade 4, Test Level-3

PUB. Percentile	Form A		Form B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	12	9	10	6
10	23	19	18	12
15	29	28	25	23
20	36	37	34	31
25	45	42	39	40
30	49	47	48	48
35	54	55	56	56
40	58	59	60	61
45	63	65	67	65
50	70	68	69	69
55	72	72	73	73
60	77	78	77	75
65	80	81	79	79
70	84	83	82	83
75	87	86	85	86
80	90	89	89	89
85	93	92	92	92
90	96	95	95	95
95	97	97	98	97

Table 48. Total Reading Scores: Basic, Grade 5, Test Level-3

PUB. Percentile	Form A		Form B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	10	9	6	8
10	19	21	13	19
15	28	38	24	31
20	37	38	34	40
25	41	44	41	45
30	47	51	48	51
35	54	55	53	55
40	58	61	57	61
45	63	66	64	66
50	68	71	67	70
55	71	75	71	72
60	74	79	75	77
65	79	83	78	81
70	83	85	81	84
75	86	87	85	87
80	88	90	89	89
85	91	93	91	92
90	93	95	94	95
95	96	97	97	97

Table 49. Total Math Scores: Basic, Grade 3, Test Level-2

PUB. Percentile	Form A		Form B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	8	9	10	9
10	17	15	16	15
15	23	20	23	23
20	31	25	28	26
25	37	29	32	33
30	43	34	37	36
35	49	40	44	43
40	54	45	47	48
45	56	50	53	51
50	62	54	58	57
55	66	59	62	60
60	71	63	67	65
65	76	67	73	69
70	80	69	79	74
75	84	72	83	77
80	86	75	88	80
85	90	80	92	84
90	93	84	94	88
95	95	90	97	92

Table 50. Total Math Scores: Basic, Grade 4, Test Level-2

PUB. Percentile	Form A		Form B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	7	Level 2	7	Level 2
10	16	not used	14	not used
15	23	in spring	21	in spring
20	28		30	
25	36		35	
30	42		43	
35	47		51	
40	52		53	
45	56		59	
50	60		64	
55	66		66	
60	66		71	
65	71		74	
70	74		76	
75	77		80	
80	80		84	
85	84		86	
90	88		90	
95	88		93	

Table 51. Total Math Scores: Basic, Grade 4, Test Level-3

PUB. Percentile	Form A		Form B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	11	12	13	12
10	19	20	23	21
15	24	28	30	27
20	31	33	34	33
25	38	40	42	39
30	45	42	45	45
35	51	47	49	48
40	54	49	56	51
45	57	56	60	57
50	62	57	64	60
55	64	60	67	64
60	69	64	70	69
65	73	69	74	71
70	76	71	77	75
75	80	73	79	79
80	82	76	85	81
85	86	80	88	84
90	89	83	91	87
95	92	88	95	92

Table 52. Total Math Scores: Basic, Grade 5, Test Level-3

PUB. Percentile	Form A		Form B	
	Fall ESAA %ILE	Spring ESAA %ILE	Fall ESAA %ILE	Spring ESAA %ILE
5	10	10	10	12
10	17	19	19	22
15	26	24	25	29
20	30	32	32	34
25	37	39	38	41
30	41	43	41	47
35	45	47	47	49
40	49	52	52	55
45	51	56	55	60
50	56	59	61	65
55	58	62	63	70
60	63	66	68	72
65	68	68	70	76
70	70	70	75	78
75	72	74	80	81
80	77	79	82	85
85	81	81	85	88
90	84	85	89	91
95	88	91	92	95

Table 53. Distribution of Gain Scores
Reading, Grade 3

ESAA Scale Gain	Percentile Rank	Publisher's Scale Gain	Percentile Rank
No Gain	22	No Gain	24
5	25	5	32
10	29	10	32
15	32	15	33
20	36	20	55
25	40	25	56
30	44	30	56
35	48	35	75
40	52	40	76
45	57	45	76
50	60	50	88
55	64	55	88
60	68	60	87
65	71	65	95
70	74	70	95
75	77	75	95
80	80	80	96
85	83	85	97
90	84	90	98
95	86	95	98
100	88	100 or more	99
105	90		
110	91		
115	92		
120	93		
125	94		
130	95		
135	95		
140	96		
145	96		
150 or more	97		

Table 54. Distribution of Gain Scores
Reading, Grade 4

ESAA Scale Gain	Percentile Rank	Publisher's Scale Gain	Percentile Rank
No Gain	29	No Gain	32
5	32	5	40
10	36	10	40
15	39	15	43
20	42	20	56
25	45	25	56
30	49	30	57
35	52	35	71
40	56	40	71
45	59	45	72
50	62	50	83
55	65	55	83
60	68	60	83
65	71	65	90
70	73	70	90
75	75	75	91
80	76	80	93
85	79	85	95
90	81	90	95
95	83	95	95
100	85	100 or more	97
105	86		
110	88		
115	89		
120	90		
125	91		
130	91		
135	92		
140	93		
145	94		
150 or more	95		

Table 55. Distribution of Gain Scores
Reading, Grade 5

ESAA Scale Gain	Percentile Rank	Publisher's Scale Gain	Percentile Rank
No Gain	32	No Gain	33
5	36	5	41
10	40	10	41
15	44	15	50
20	48	20	60
25	52	25	60
30	57	30	63
35	60	35	77
40	65	40	77
45	68	45	77
50	71	50	85
55	75	55	88
60	78	60	88
65	80	65	92
70	82	70	95
75	85	75	95
80	86	80	96
85	88	85	97
90	89	90	97
95	91	95	97
100	92	100 or more	99
105	93		
110	94		
115	94		
120	95		
125	96		
130	96		
135	96		
140	97		
145	97		
150 or more	98		

Table 56. Distribution of Gain Scores
Math, Grade 3

ESAA Scale Gain	Percentile Rank	Publisher's Scale Gain	Percentile Rank
No Gain	13	No Gain	15
5	15	5	22
10	17	10	22
15	19	15	23
20	21	20	41
25	24	25	42
30	27	30	42
35	30	35	62
40	33	40	63
45	37	45	63
50	40	50	79
55	43	55	80
60	47	60	80
65	50	65	87
70	54	70	91
75	58	75	91
80	61	80	93
85	65	85	96
90	68	90	96
95	71	95	96
100	74	100 or more	98
105	76		
110	79		
115	82		
120	84		
125	86		
130	88		
135	89		
140	91		
145	92		
150 or more	94		

Table 57. Distribution of Gain Scores
Math, Grade 4

ESAA Scale Gain	Percentile Rank	Publisher's Scale Gain	Percentile Rank
No Gain	23	No Gain	26
5	25	5	34
10	28	10	34
15	31	15	40
20	34	20	53
25	37	25	53
30	40	30	55
35	43	35	71
40	47	40	72
45	50	45	72
50	53	50	85
55	56	55	85
60	60	60	85
65	63	65	92
70	66	70	92
75	69	75	92
80	73	80	94
85	75	85	96
90	77	95	96
95	80	95	96
100	82	100 or more	98
105	84		
110	85		
115	87		
120	89		
125	90		
130	91		
135	92		
140	93		
145	94		
150 or more	94		

Table 58. Distribution of Gain Scores
Math, Grade 5

ESAA Scale Gain	Percentile Rank	Publisher's Scale Gain	Percentile Rank
No Gain	28	No Gain	30
5	32	5	38
10	35	12	38
15	38	15	45
20	43	20	57
25	46	25	57
30	49	30	62
35	53	35	75
40	57	40	75
45	60	45	77
50	63	50	86
55	66	55	86
60	70	60	86
65	73	65	93
70	76	70	93
75	78	75	93
80	81	80	95
85	83	85	96
90	84	90	96
95	86	95	97
100	88	100 or more	98
105	89		
110	91		
115	92		
120	93		
125	94		
130	95		
135	95		
140	96		
145	96		
150 or more	97		

Table 59. Distribution of Gain Scores: Secondary Sample

Publisher's Scale Gain	Grade 10		Grade 11		Grade 12	
	Reading Percentile Rank	Math Percentile Rank	Reading Percentile Rank	Math Percentile Rank	Reading Percentile Rank	Math Percentile Rank
No Gain	46	39	49	44	53	47
5	62	44	65	49	68	52
10	76	49	78	54	81	57
15	85	54	85	59	88	62
20	91	57	91	62	92	65
25	95	60	94	65	94	68
30	97	63	96	68	95	72
35	98	66	96	71	96	74
40	98	69	97	74	96	76
45	99	72	97	77	97	78
50	99	75	98	80	97	81
55	99	78	98	83	97	83
60	99	81	98	86	97	86
65	99	84	98	89	97	91
70	99	87	98	92	97	92
75	99	90	98	93	97	93
80	99	93	98	95	97	94
85	99	94	98	96	98	94
90	99	95	98	97	98	95
95	99	96	98	97	98	96
100 or more	99	97	98	97	98	96

Table 60. Approximate Comparison of Selected ESAA Percentiles With Anchor Test Percentiles and Corresponding Publisher's Percentile Ranks on Total Reading Scores: Grade 4

ESAA	ANCH	CAT	CTBS	GMT	ITBS	MAT	SAT	SRA	STEP
99	95	96	95	94	94	93	95	95	95
95	85	85	84	83	83	82	87	85	82
90	75	74	74	75	73	71	71	73	71
85	68	65	67	68	65	63	62	65	63
80	60	57	59	60	57	55	54	58	56
75	52	50	53	54	50	47	45	50	49
70	45	45	46	48	42	39	36	43	42
65	37	38	39	40	35	34	30	35	34
60	30	33	31	32	28	28	23	28	27
55	27	28	28	28	26	24	20	24	23
50	23	24	25	25	23	20	17	21	20
45	20	21	23	22	20	16	14	18	17
40	17	18	19	18	16	13	13	15	14
35	13	15	15	13	12	10	11	12	11
30	10	12	12	8	8	7	9	10	8
25	8	10	11	7	7	6	8	8	6
20	6	7	8	6	6	4	6	6	5
15	5	6	7	4	4	2	5	5	4
10	4	4	4	3	3	1	3	3	3
5	1	1	1	2	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1

Table 61. Approximate Comparison of Selected ESAA Percentiles and Corresponding Publisher's Percentile Ranks on Total Reading Scores: Grade 5

ESAA	ANCH	CAT	CTBS	GMT	ITBS	MAT	SAT	SRA	STEP
99	98	98	98	97	97	98	96	97	97
95	87	87	85	82	84	84	80	85	84
90	76	74	73	71	71	75	65	75	71
85	67	66	64	62	62	67	54	66	62
80	58	58	56	54	53	60	46	57	53
75	50	51	48	47	44	48	38	48	45
70	44	44	42	41	39	45	32	43	41
65	38	38	37	35	34	36	27	36	33
60	33	33	32	29	27	31	22	30	28
55	26	28	27	24	21	26	17	24	23
50	23	25	24	21	18	22	15	21	20
45	20	22	21	19	16	18	13	19	17
40	18	19	18	16	14	16	11	16	14
35	15	16	16	13	12	14	10	14	12
30	13	14	13	10	9	12	8	11	10
25	10	11	11	8	7	10	7	9	8
20	7	8	9	5	5	7	5	6	6
15	5	6	7	3	3	4	4	4	5
10	4	4	5	2	2	2	2	2	3
5	2	2	3	1	1	1	1	1	2
1	1	1	1	1	1	1	1	1	1

3. COMPARISONS OF ACHIEVEMENT GROWTH OBTAINED FROM CROSS-SECTIONAL AND LONGITUDINAL SAMPLES IN THE ESAA EVALUATION

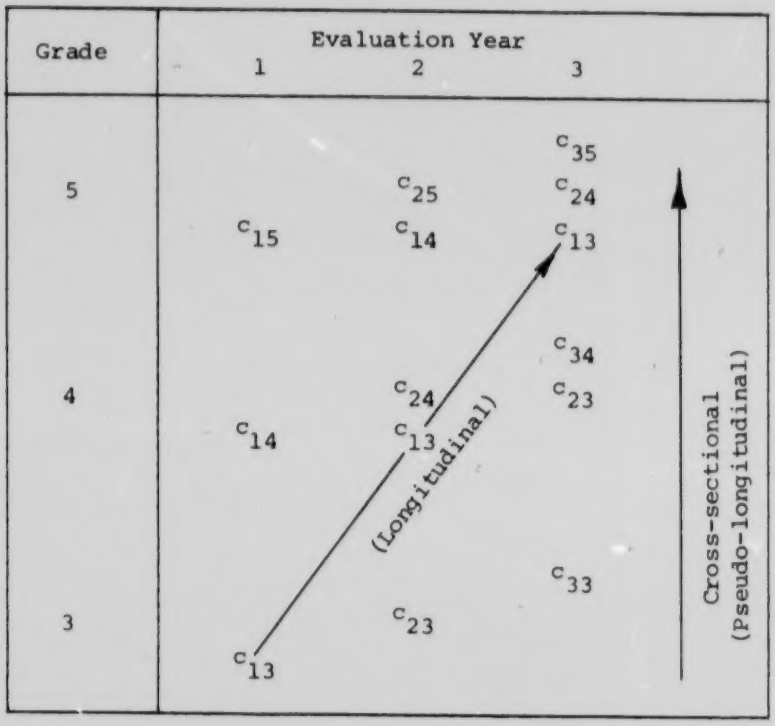
This section of the report compares estimates of student achievement growth in reading and mathematics obtained from different sampling designs in order to examine possible effects of those design differences on apparent growth. The measurement of achievement change is generally described within the context of cross-sectional or longitudinal designs. The term 'longitudinal design' encompasses any design in which repeated measures are obtained over a specified time period from the same or comparable populations; for example, from the 3rd grade one year and from the 4th grade the following year. A matched longitudinal design utilizes data from only those students for whom all measures are available--in this case, scores from both the 3rd-and-4th grade testings. An unmatched longitudinal design includes all students present at each time of measurement regardless of whether they were in the sample at any prior or subsequent measurement point; thus, the sample in a matched longitudinal design would be a subset of the unmatched longitudinal sample. A cross-sectional study collects data from students in different grades at a single point in time; for example, in the Survey of Equality of Educational Opportunity (Coleman, 1966), a national sample of students in the 1st, 3rd, 6th, 9th, and 12th grades was tested at the same time. Achievement estimates from students in the various grades can be combined to yield what we here describe as psuedo-longitudinal estimates of achievement growth.

Despite possible biases introduced in the "true" or matched longitudinal design, such as attrition and retest effect, there is in educational research a general acceptance of the superiority of data collected over an extended time period, compared with data collected at a single time point (Campbell and Stanley, 1963). Given the difficulty of collecting data over an extended period of time and the expense of long-term experiments, it is relevant to determine the extent to which estimates of academic achievement derived from the alternative designs differ.

a. Method

Data obtained during the three-year evaluation of the Emergency School Aid Act (ESAA) provided an opportunity to estimate growth in achievement using alternative designs. A simplified schematic representation of the ESAA sample is provided in Figure 17, c_{ij} is a cohort identifier in which i is the year of entry of the sample student into the study and j is the grade of the student at entry. A pretest and posttest were given within each year. The vertical (between-rows) comparisons are cross-sectional; longitudinal comparisons are on the diagonals. As can be seen in the figure, cross-sectional estimates were obtained for each

Figure 17. Cross-Sectional and Longitudinal Designs



of three years; there was one three-year longitudinal sample (c_{13}) and several two-year longitudinal samples.*

Results from forty schools that participated in the evaluation for all three years were used in this analysis. The schools consisted of twenty matched pairs, with one school from each pair receiving ESAA Basic funds and the other school, though eligible for funds, being denied those funds. The analysis was restricted to this particular subset of schools in order to avoid the necessity of estimating the effects of sample school attrition and supplementation. It was felt that this more nearly approximated the usual short-term experiment in which sample mobility is confined to students. The schools were drawn from a population of schools with substantial numbers of disadvantaged and minority students; in most cases, the districts in which the schools were located had previously desegregated.

A random sample of eligible 3rd-, 4th-and-5th grade students within each school was selected at the start of the study.** Students entering the sample were followed through the grade range of interest and automatically dropped after the 5th grade. New student samples were drawn at the 3rd grade in the second and third years, and students were added to grades 4 and 5 to offset attrition. There were approximately 1700 students in each grade in each of the three years. Of the 3rd-grade students entering the sample in the first year, about 900 remained in the sample the full three years and had test scores available for the pretest and posttest each year.

---

*On the three yearly ESAA reports (Coulson, 1975-1977), within-year changes between the pretest and posttest were described in relation to each year's cross-sectional sample. This was in order to differentiate them from results obtained from the longitudinal sample, which was restricted to those students who were in the study from its start. This report, however, uses the term 'cross-sectional' in its technical sense; that is, to refer to data collected at a single point in time.

**Eligible students were those who participated in the ESAA-funded programs in schools receiving such funds or who were eligible to participate in such programs in nonfunded schools. Additional details of the study are provided in the overview of the three years of ESAA evaluation (Coulson, 1977).

Students were administered the California Achievement Test (CAT) in the fall and spring of school years 1973-1974, 1974-1975, and 1975-1976. They were randomly assigned to one of two alternate forms (A or B) of the appropriate test-level (2 or 3) at the fall testing, and given the other form at the spring testing. Because of problems such as the disparity in difficulty of the alternate forms and levels of the test, the publisher's scale scores could not be used to adequately describe achievement growth trends; therefore, the tests were restandardized using the ESAA sample as the norm group, (Doherty, 1977). The revised ESAA scale scores were used for all estimates; thus, we are assuming that errors in equating the various forms and levels were negligible.

b. Cross-Sectional and Longitudinal Comparisons

Students were tested in the fall and spring of each of three study years, so that there were six cross-sectional estimates of achievement. Each estimate was based upon all the students tested at that point in time. These estimates are shown graphically in Figure 18 by the three broken lines in each box. The average ESAA reading and math scale scores for the 3rd, 4th, and 5th grades in each of the three years are shown separately for the fall and spring testings. The corresponding means, standard deviations, and sample sizes are provided in Table 62.

As can be seen, the achievement estimates for the first-year (1973-1974) cross-sectional samples were consistently lower than those for the second year (1974-1975) or the third year (1975-1976) cross-sectional samples, particularly at the fall testing.* Figure 19 suggests two possible reasons for this difference.**

*Because of the large sample sizes, a scale score difference of about five points (about one test item) was sufficient for a statistically significant difference between two groups of students; therefore, almost all differences discussed were statistically significant, although they are not necessarily meaningfully different.

**The figure is based upon students who had complete data from their time of entry into the sample until the conclusion of their participation. That is, the 5th-grade students entering the sample each year had both a pretest and a posttest. The 4th-grade students entering the sample in the first year or second year had a pretest and posttest for both the 4th and 5th grades; however, the 4th-grade students entering the sample in the third year had just the pretest and posttest for the 4th grade.

Figure 18. Longitudinal and Cross-Sectional Comparisons of Reading and Math Mean Achievement: Fall and Spring Testings

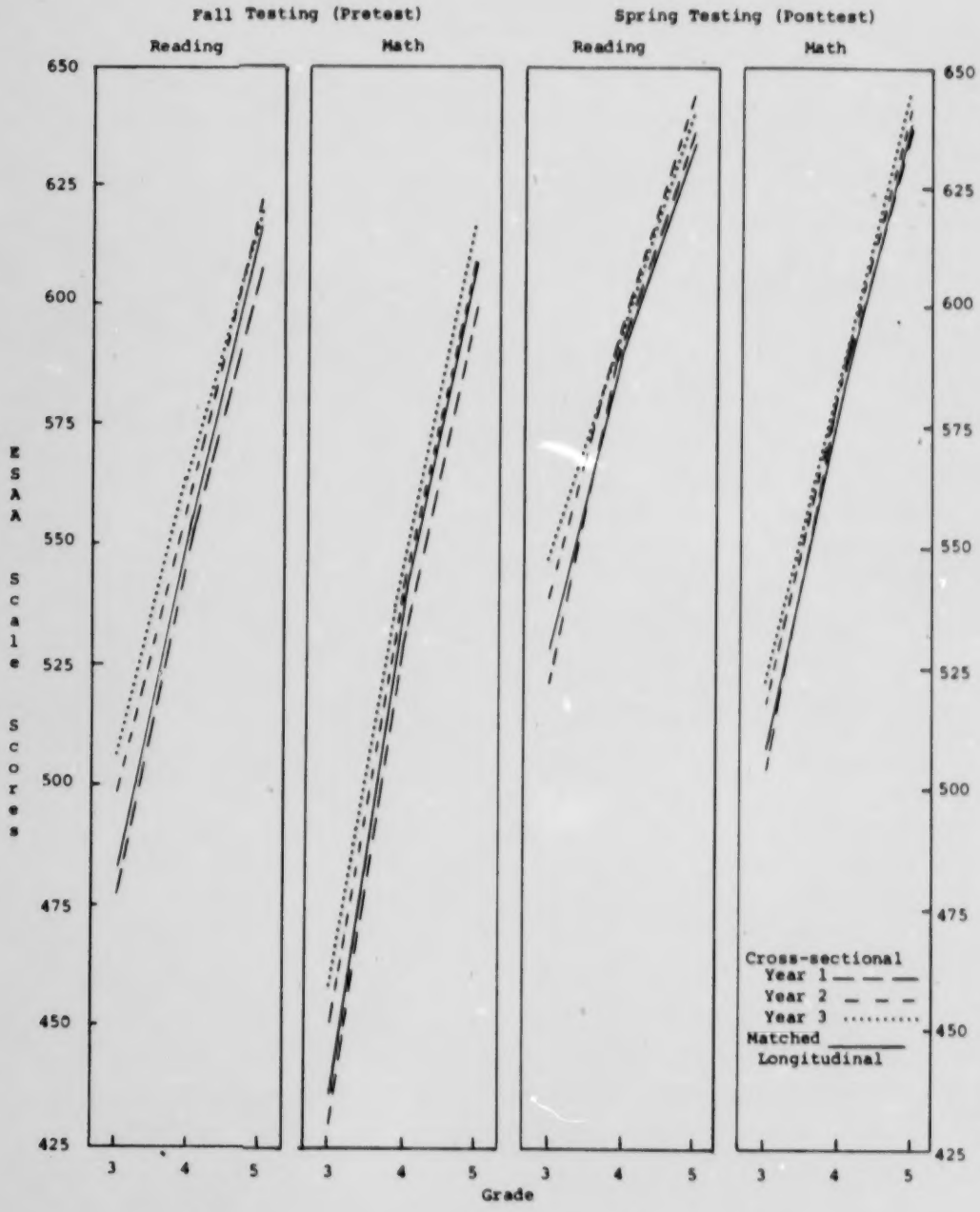


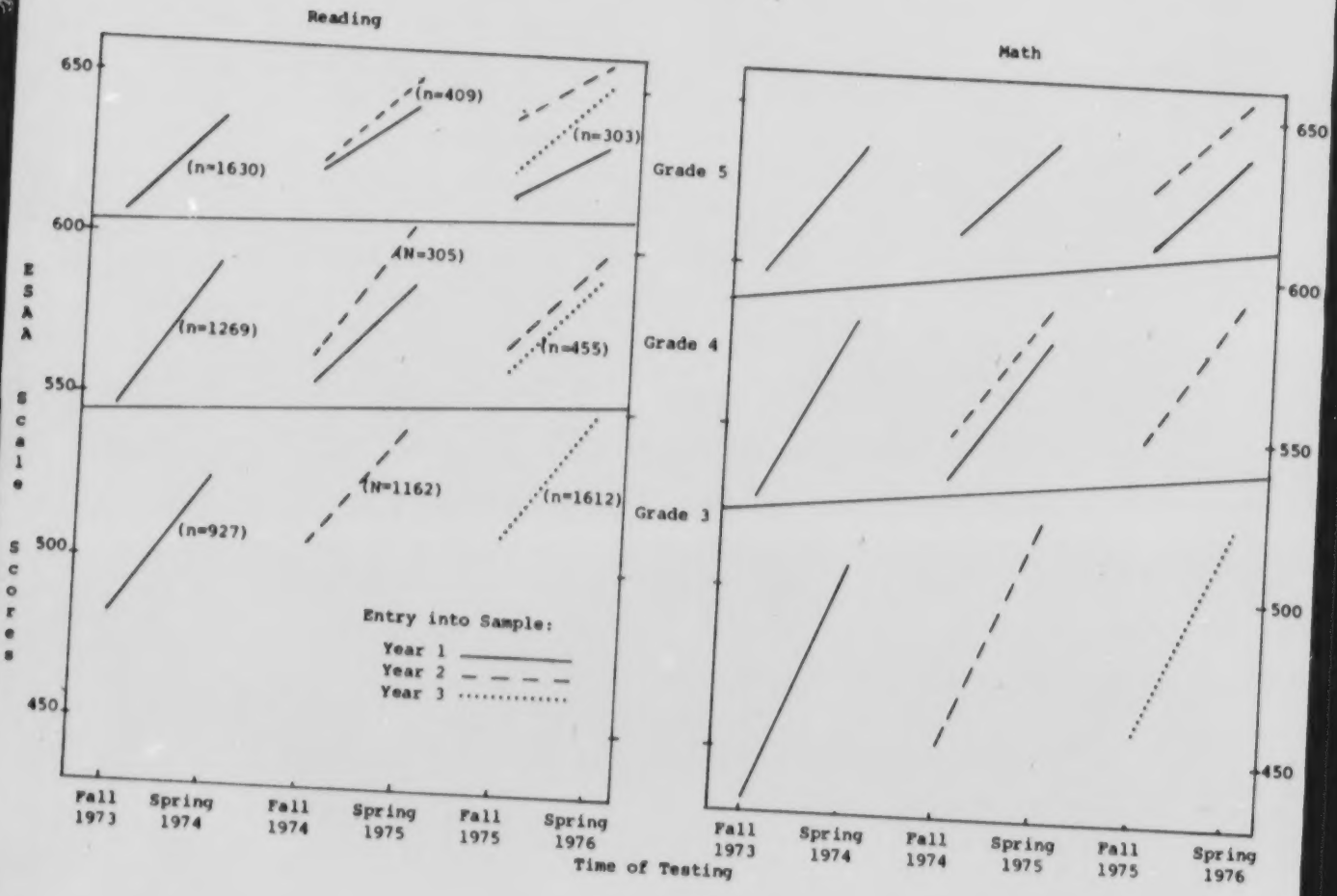
Table 62. Longitudinal and Cross-Sectional Comparisons of Reading and Math Achievement

Grade	Reading					Math				
	Cross-sectional Year			Longitudinal*		Cross-sectional Year			Longitudinal	
	1	2	3	Matched	Unmatched	1	2	3	Matched	Unmatched
3. Pretest $\bar{x}$	477.5	500.1	507.8	482.7	477.5	429.8	447.9	456.2	432.5	429.8
s.d.	112.0	107.8	103.8	111.2		79.4	80.1	77.6	78.1	
n	1835	1773	1765	927		**				
Posttest $\bar{x}$	522.7	539.4	548.0	525.9	522.7	502.9	517.8	521.7	507.4	502.9
s.d.	111.1	109.1	103.8	106.1		95.8	96.2	92.2	91.7	
n	1697	1625	1615							
4. Pretest $\bar{x}$	549.9	560.1	565.2	554.5	560.1	528.5	539.0	544.4	536.8	539.0
s.d.	101.2	104.4	105.7	104.8		91.5	94.8	92.0	94.8	
n	1873	1864	1761							
Posttest $\bar{x}$	591.6	596.1	595.2	589.2	596.1	582.7	583.1	585.0	579.8	583.1
s.d.	97.9	103.1	101.4	101.6		100.0	104.0	103.3	105.9	
n	1756	1808	1707							
5. Pretest $\bar{x}$	607.9	622.6	621.5	616.4	621.5	598.8	609.7	615.2	609.3	615.2
s.d.	104.1	98.8	102.4	103.4		99.5	98.3	104.5	105.9	
n	1766	1795	1673							
Posttest $\bar{x}$	635.1	644.7	641.6	632.9	641.6	637.3	641.7	646.7	639.3	646.7
s.d.	103.0	98.2	102.9	104.5		106.7	106.4	109.4	109.9	
n	1689	1746	1651							

*The unmatched longitudinal samples are the same as successive cross-sectional samples

**N's for reading and math are the same

Figure 19. Mean Achievement (By Year and Grade of Entry into Sample) of One-, Two-, and Three-Year Longitudinal Samples



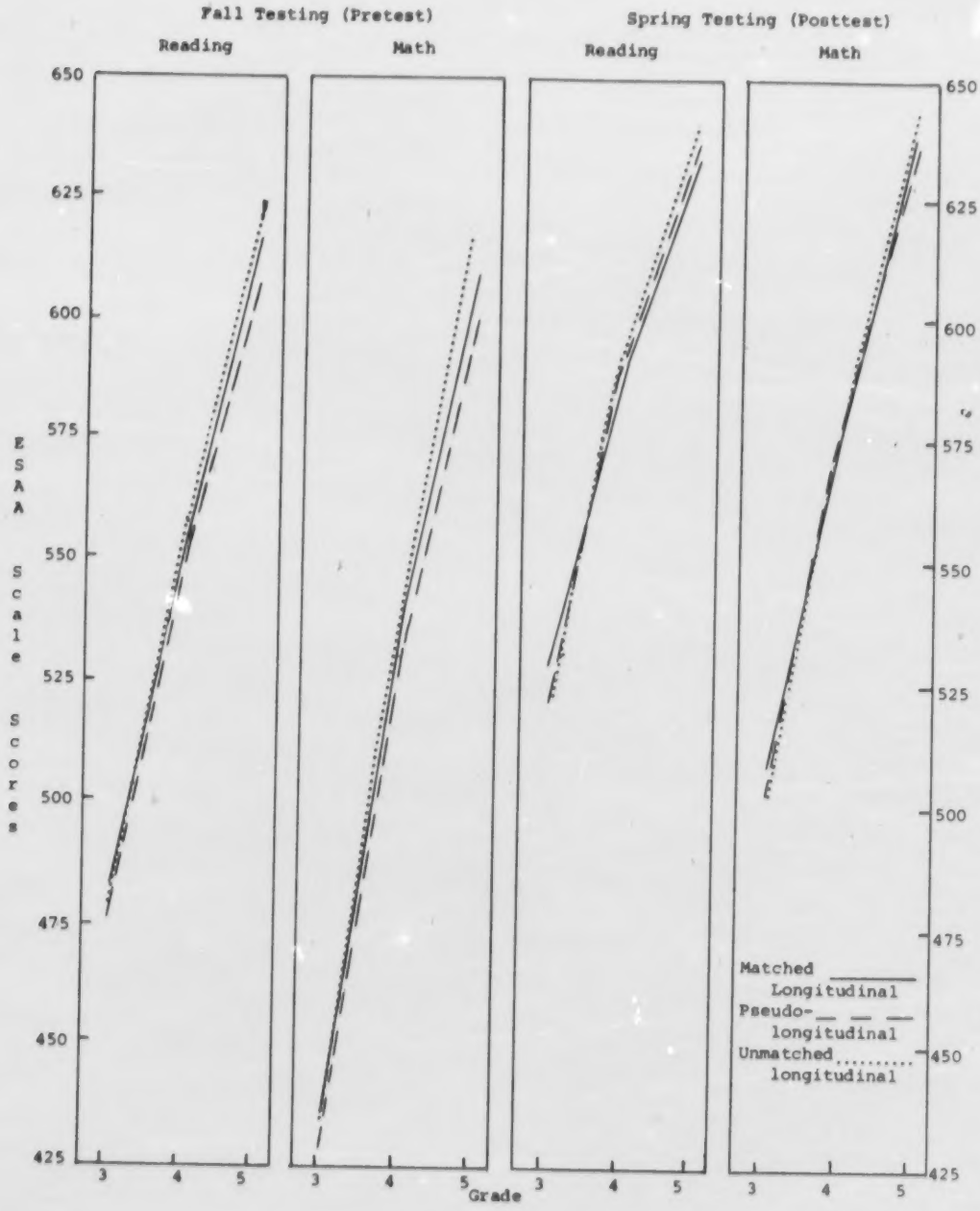
First, the performance of students entering the sample in the second year of the study (shown by the broken line) was superior to the performance of the students entering the first year (shown by the solid line). For example, the reading score for the first-year 3rd-grade students was 482.7, constrasted with a score of 506.2 for the 3rd-grade students entering the sample in the second year. Second, the performance of students remaining in the sample in the second year was superior to the performance of the students in the same grade the previous year. For example, during the first year, 5th-grade students scored an average of 607.9 on the reading pretest. This constrasted with an average score of 621.8 achieved during the second year by 5th-grade students who had been in the 4th-grade sample the first year.

For the sake of visual clarity, estimates of achievement growth from the matched longitudinal and pseudo-longitudinal (first-year cross-sectional) students are repeated in Figure 20 along with estimates based on the unmatched longitudinal students. The unmatched longitudinal sample included all students present in the sample at the same time and grade level as the matched longitudinal sample, i.e., the first-year 3rd-grade, the second-year 4th-grade, and the third-year 5th-grade students.* Estimates derived from the matched and ummatched longitudinal samples were not independent, since, if there were little movement of students into and out of the sample, the matched and unmatched groups would be similar. In this analysis, the matched longitudinal sample comprised about half the unmatched sample.

As can be seen in Figure 20, the unmatched longitudinal sample showed greater growth in both reading and math achievement than did either the pseudo-longitudinal (cross-sectional) or matched longitudinal samples. This would appear to be due, as was the difference between the first-year cross-sectional sample and subsequent cross-sectional samples, to the superior performance of students added to the sample after the first year.

*Had this analysis not been restricted to schools remaining in the study all three years, the unmatched longitudinal sample would not be equivalent to the sequential cross-sectional samples. The cross-sectional samples under such circumstances would have included students in schools that left the sample prior to the conclusion of the experiment as well as students in replacement schools.

Figure 20. Matched, Unmatched, and Pseudo-Longitudinal Comparisons of Reading and Math Mean Achievement: Fall and Spring Testings



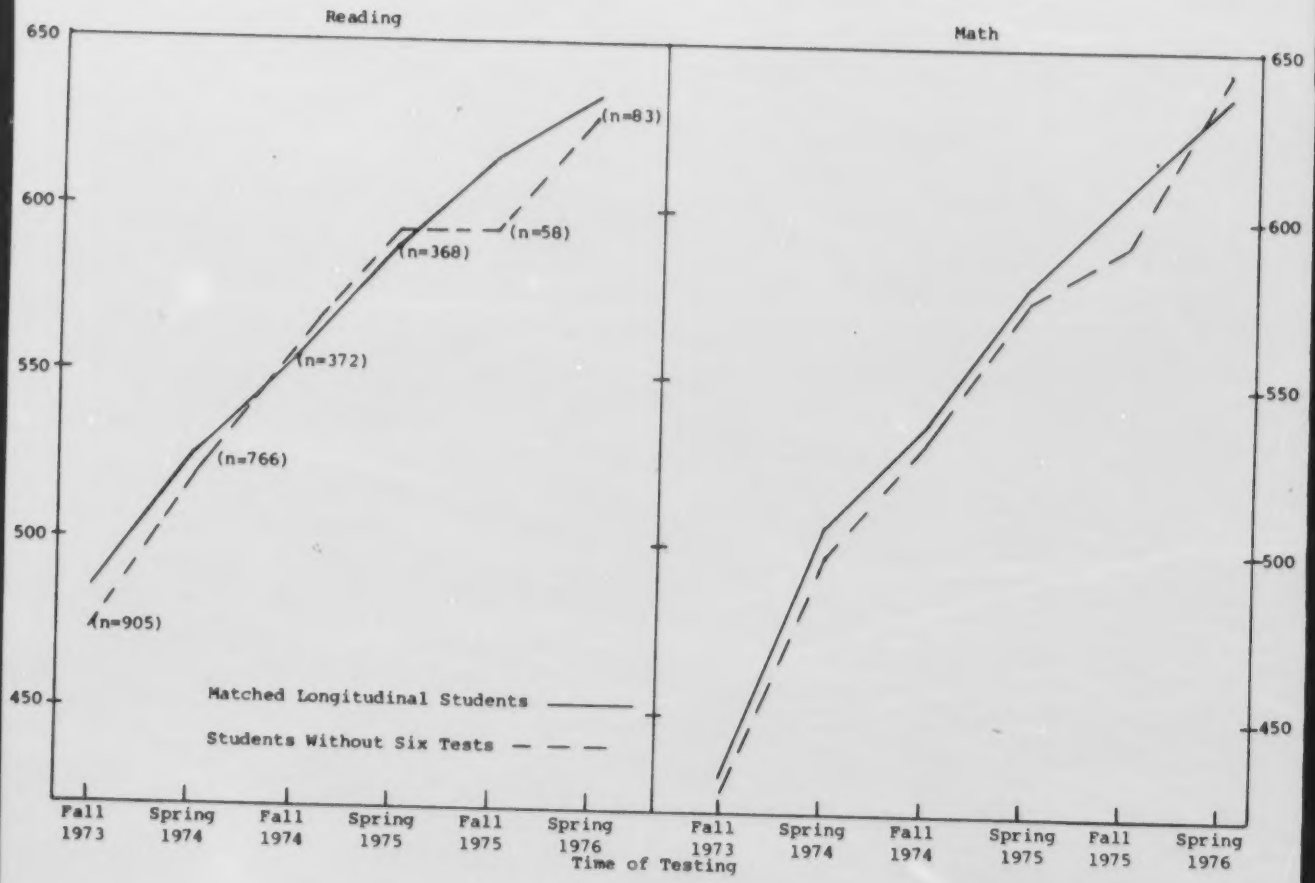
A comparison of the matched longitudinal estimates (the solid line in Figure 24) with the pseudo-longitudinal estimates based upon the first-year cross-sectional data (the broken line) resulted in slightly different interpretations for the fall and spring testings. Although 3rd-grade achievement estimates were very similar for the matched longitudinal and first-year cross-sectional samples for both the pretest and posttest, data from the 4th and 5th grades at the fall testing resulted in matched longitudinal estimates of achievement growth that were slightly greater than the pseudo-longitudinal estimates. However, the spring matched longitudinal and pseudo-longitudinal estimates were almost identical.

One might suspect from the fall testing that differential attrition has occurred, i.e., that students who remained in the sample and had complete data performed better than those who left the sample or missed some tests. However, a comparison of the longitudinal students with the first-year 3rd-grade students who did not take all six tests (see Figure 21) does not bear this out. The solid line represents those students who were in the 3rd-grade sample at the first-year fall testing but left the sample at varying points, or who did not take all six tests. The number of such students is shown in parentheses at each test point; there were 927 longitudinal students at each test point. Although the matched longitudinal students scored somewhat higher in math, those with incomplete data were superior at the final testing. In reading, however, differences between the two groups of students were not consistent. Thus, attrition appears to have little effect on measurement of change in the ESAA sample.

c. Summary and Conclusions

Are different estimates of reading and mathematics growth obtained when different sampling strategies are used? In the samples of students drawn at the same time from a finite population of schools, differences in estimates of growth from the matched longitudinal and cross-sectional samples were relatively small, although the matched longitudinal sample revealed slightly greater achievement growth when compared with pseudo-longitudinal growth projections based on the first-year fall cross-sectional sample. Student attrition did not appear to bias the estimates in the ESAA sample, although other studies have called attention to the problem of selective attrition (Hilton and Patrick, 1970).

Figure 21. Achievement Means for Longitudinal Students and Students Without Complete Data



When samples of different students were measured at different times, as in the unmatched longitudinal sample (Figure 20), the growth estimates were greater than those from the matched longitudinal sample and also greater than the pseudo-longitudinal estimates from the cross-sectional samples, because of the superior performance of students after the first year. However, under other circumstances the differences might have gone in either direction, depending upon the relationships among the students' achievement at each time of testing.

Thus, it would seem that the least reliable estimates of achievement growth are those in which both the time and students are different, as in the unmatched longitudinal sample. When just one factor varies--that is, when the students in all grades are the same but the time changes, as in the matched longitudinal sample, or when the time is the same but the students in each grade are different, as in the cross-sectional sample--then the estimates are relatively similar.

B. REANALYSES OF DATA USING DERIVED INDICES AND COMPOSITE VARIABLES

This section describes activities directed toward the development of new composites and indices from the previously collected program, staff, and student data, and presents the results of additional analyses exploring relationships among some of the original and newly derived measures. The first subsection discusses the derivation of new predictor variables from the data on student and staff attitudes and from questionnaire data on the sample schools and programs. The second subsection describes efforts to derive new outcome measures that might shed greater light on the students' learning strengths and weaknesses, and perhaps reveal stronger relationships with certain predictor variables. Third, several different approaches to reanalysis of the ESAA program impact data are discussed, and results of those reanalyses are summarized. Finally, the methodology and results of supplemental analyses--designed to identify student and program factors influencing achievement gains, i.e., relational analyses--are discussed.

1. DERIVATION OF NEW COMPOSITE VARIABLES CHARACTERIZING PROGRAMS AND STUDENTS

In the original ESAA analyses, as documented in the three yearly evaluation reports, many of the variables descriptive of schools, programs, and staff were represented by only a single questionnaire item. While such limited variables have the advantage of being easily constructed and readily interpretable, they tend to give a segmented, disjointed view of the schools being characterized, and of the staff and students in those schools. Thus, it can be difficult from a simple collection of such one-item variables to get a synthesized view or profile of the different types of programs being contrasted. For this reason, one of the goals of the additional analyses discussed here was to reexamine some of the questionnaire data and to derive new composite variables that might provide a better global picture of some major clusters or profiles of sample schools and their associated program characteristics. A related goal, discussed later, was to explore the relationships among these new variables and measures of program outcomes (achievement gains) to determine whether those relationships might alter, confirm, or further explicate conclusions previously drawn about the relative effectiveness of different program approaches.

a. New Attitude and Exposure Indices

The first set of variables examined related to teacher and student attitudes and to student exposure to academic activities. Relevant questionnaire items were recoded so that the direction of response coding was consistent within and across questionnaires, e.g., so that a high value would always mean a favorable attitude. Missing values were imputed using median or modal response categories. Intermediate indices were then constructed and, finally, summary indices were created from combinations of the intermediate indices, as described in the following paragraphs.

The student attitude index was constructed in two stages. First, separate indices of locus of control (LOC), belongingness (BLNG), and attitude towards school (ATTD) were constructed as weighted sums of items. Second, an overall index (STATD) was calculated as the sum of these intermediate indices. An index for teacher attitude was similarly constructed. Intermediate indices of attitude towards teaching, and feelings of frustration were constructed as weighted sums of appropriate items on the questionnaires. The overall teacher attitude index (TCHATD) was the sum of these intermediate indices.

Three indices of academic exposure were computed; these summarized total school exposure, total exposure to mathematics, and total exposure to reading. Total school exposure (TOTSCH) was coded in terms of the number of hours of school attended by the student per year. This was based upon data concerning the number of school days per year and hours per day, adjusted for the number of absences for each student. The math exposure index (TOTMATH) and reading exposure index (TOTREAD) represent, respectively, the number of hours per year of math and reading instruction received by the student.

When the academic exposure index scores of students in treatment and control schools were compared, an interesting pattern emerged. As can be seen in Table 63, students in treatment schools spent less total time in school than did students in control schools; further, the treatment students' exposure to reading and math instruction was significantly greater than that of control group students in the 3rd and 5th grades, but about the same in the 4th grade.

The relationship of these differences to reading and math achievement gains is discussed in a later section of this report.

Table 63. Treatment-Control Comparisons of Academic Exposure

Type Of Exposure (hours per year)	Grade 3 Treatment	Grade 3 Control	Grade 4 Treatment	Grade 4 Control	Grade 5 Treatment	Grade 5 Control
Total	913.0	927.9*	932.0	964.4*	936.4	960.6*
Reading	352.2	330.0*	336.3	333.1	321.8	294.9*
Math	181.2	166.5*	182.2	182.2	184.2	175.7*

*The difference between the treatment and control groups was statistically significant at $p < .01$.

b. Program Characteristic Indices

The reexamination of school and program variables had two objectives: first, to determine whether program and school characteristics exhibited patterns permitting description of distinct types of schools and, second, to determine whether such school (and/or classroom) differences related to differential gains in student reading and math achievement.

The reanalysis described here examined groupings of variables, rather than separate item-variables, to determine whether meaningful patterns of classroom and school characteristics existed. Such analyses were confined to third-year data, and to data from the Basic elementary sample, which was the largest and most stable of the three samples.

To reduce the large number of variables to a more manageable number for use in data processing and analysis, indices were developed. Emphasis was placed on finding indices that would reflect meaningful dimensions of the school or classroom and that would differentiate schools from one another and/or relate to changes in students' achievement levels. Therefore, the initial selection of items for reanalysis was based upon their potential relevance to program areas

of interest, their face validity, and their possible relation to achievement. The items were selected from two sources: principal responses to questions about school and program characteristics, and teacher responses concerning school, program, and classroom characteristics. Missing items were assigned the modal response of the sample, but no imputations were made for nonrespondents (missing questionnaires).

Two methods were used to develop the indices. First, variables that had reasonably high correlations with each other and seemed conceptually related were subjected to factor analysis.* The variables with high loadings (generally over .35) on factors that could be meaningfully interpreted were standardized, assigned unit weights, and summed. Secondly, variables that were not necessarily highly correlated, but were felt on rational grounds to represent different aspects of the same dimension, were standardized and summed. An example of this would be the program emphasis variables in which supplementary dollar expenditures for remedial programs and resource allocation costs were not highly correlated, yet were felt to represent different components of program emphasis. All variables were standardized to a mean of 100 and a standard deviation of 10, and the mean score for each school was computed for all data. The resultant indices and several variables that were carried along separately are described immediately below.

School and Classroom Indices

(1) ABIL — Students are assigned to regular classes on the basis of ability and scores on standardized tests of achievement or intelligence.

(2) CLSS — Class size

(3) COUN — Counseling and guidance program emphasis. Supplementary funds used for the counseling and guidance program; the percentage of students (in sample classes) receiving group or individual counseling.

*See Harmon (1967) for a general treatment of factor analysis, and SPSS (Nie, N.H., 1975) for details of the computer program utilized.

(4) CULT — School and classroom emphasis on cultural heritage/intergroup relations programs. Supplementary funds used (including for community activities); principal and teacher participation in classroom and nonclassroom activities.

(5) DIS — The educational and economic disadvantagement of the school and sample student instructional group.

(6) DIST — District involvement in school and classroom decision-making (course content, budgeting classroom resources, hiring and firing of teachers, and classrooms texts and materials).

(7) GRP — Emphasis on small groups, use of support personnel with identified low-achieving students.

(8) IND — Individualization of activities, curriculum, and methodology.

(9) MATH — School emphasis on the math program. Supplementary funds used for remedial/compensatory mathematics; resource allocation costs for math instruction; principal and teacher participation in math curriculum development, program planning, and evaluation; teacher and aide inservice math training.

(10) NON — Majority of classrooms in school are nontraditional, extensive use of upgraded classes, flexible class scheduling, and mixing grades within classes.

(11) PAR — Parent involvement in school and classroom decision-making, regular class assignment by parent request, and principal support of parent involvement in classroom activities.

(12) PRIN — Principal involvement in school and classroom decision-making, and decision-making input relative to teacher input.

(13) RAND — Nonsystematic class assignment. Pupils are assigned to regular classes by student request, alphabetically, or by no predetermined bases.

(14) READ — School emphasis on the reading program. Supplementary funds used for remedial/compensatory reading; resource allocation costs for reading instruction; principal and teacher participation in reading curriculum development, program planning, and evaluation; teacher and aide inservice reading training.

(15) SPEC — Pupil class assignment in the regular instructional program based upon special needs, e.g., psychological problems, handicapped, etc., and teacher judgment of pupil's needs. Principal support of retraining teachers to teach low-achieving students.

(16) STUD — Student-oriented classroom activities. Independent use of self-instructional materials, learning centers, games, puzzles, models, manipulative devices, display materials, etc.

(17) SUBJM — Emphasis on math subject-matter activities and evaluation in a regular classroom setting.

(18) SUBJR — Emphasis on reading subject-matter activities and evaluation in a regular classroom setting.

(19) TCH — Teacher involvement in school and classroom decision-making.

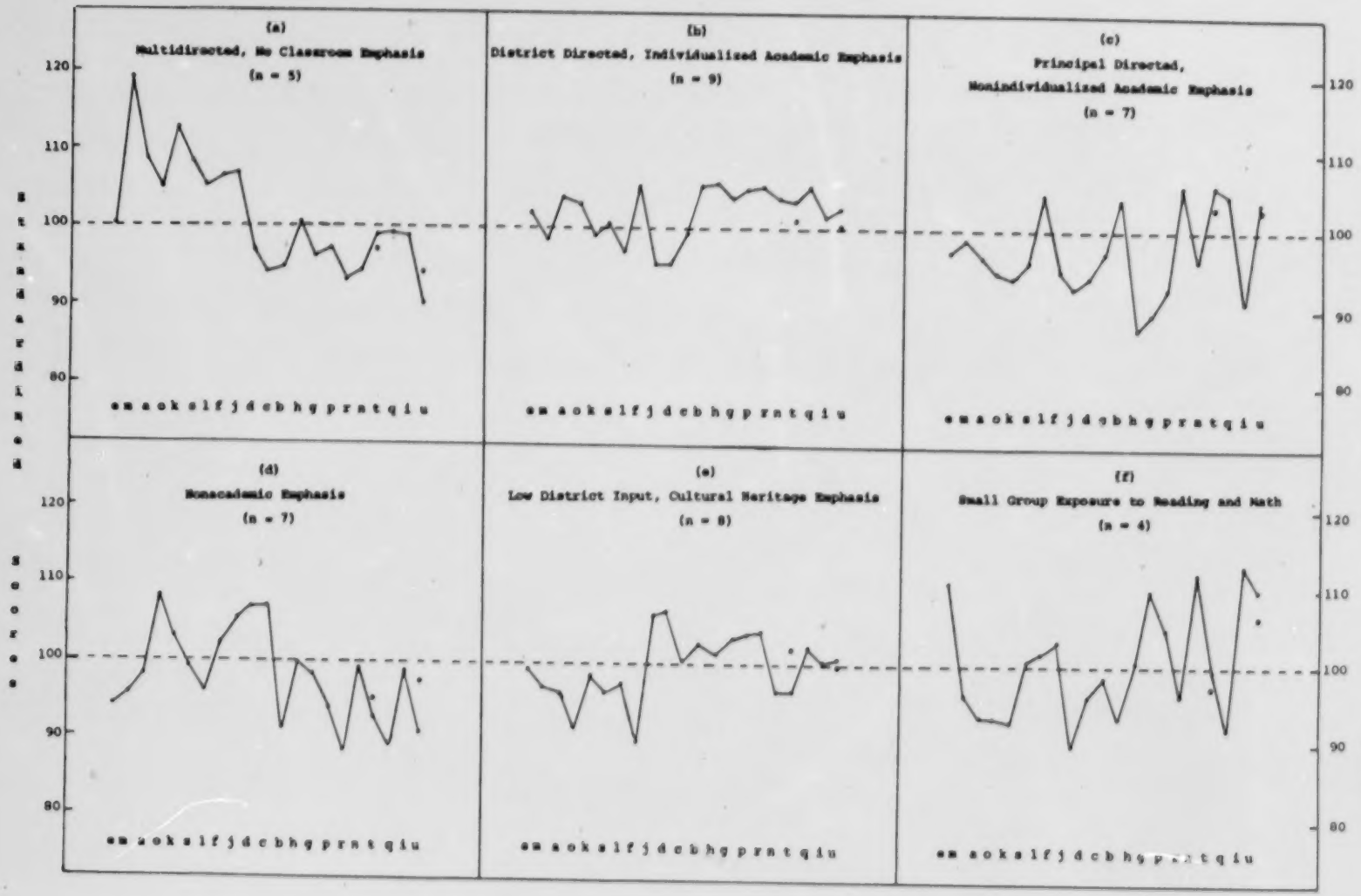
c. Description of School Profiles

The program characteristics indices and the reading and math exposure indices were subjected to hierarchical cluster analyses (Barr, 1976; Johnson, 1967). In brief, the technique is based upon grouping together close observations. Although the differences among the school values on the individual indices were not particularly large, the patterns were different and resulted in the identification of six groups of schools.

The profile for each group is shown in Figure 22. The indices are identified by the letters a through u, where a through s refer to the indices described in Section b. above, and t and u are the reading and math exposure indices. The differences among the schools were in three general areas. These were, first, the school's authority structure; that is, the participation by the district, principal, teachers, and parents in the classroom and school decision-making processes (see indices f, k, l, and s). Second, the school's program emphasis, that is, the personnel and funding resources, classroom time, etc., devoted to the program areas of reading, math, cultural heritage/intergroup relations, and counseling (indices c, d, i, n, q, r, t, u). And, third, the classroom assignment and grouping practices, such as class size, individualized instruction, assignment on the basis of special need, etc. (indices a, b, g, h, j, m, o, p).

The figures also include the average reading and math gain scores, which have been transformed to the same scale as the characteristics, for each group. They are depicted by open circles in the appropriate exposure columns (t for reading and u for math). As can be seen, reading and math gain scores generally appear

Figure 22. School Profiles



to be related to the student exposure to reading and math instruction. The relationship of gains to exposure and other program characteristics is discussed in a later section of this report.

The first group (Figure 22a) could be called multi-directed with no clear-cut classroom emphasis. These schools had a variety of class types with some student classroom assignment based on needs and ability, but most based upon non-student related methods, such as alphabetical assignment. All school-related groups were involved in decision-making regarding the school and classroom, with particularly high input by parents. There was little emphasis on reading activities, and the amount of student exposure to math was relatively low compared with other sample schools.

The second group of schools (Figure 22b) was district-directed, with emphasis upon math and reading subject matter in a large classroom environment. However, there was individualization of activities and use of student-oriented material as well as the more standard classroom subject matter.

Within the third group of schools (Figure 22c) direction was provided by the principal. There was emphasis upon reading and math in a conventional classroom setting, with little use of support personnel or individualization. Despite the classroom academic emphasis, there was a low level of school-wide focus upon the math program, in that there were low levels of supplementary expenditures, resource allocations, personnel training, etc.

The fourth group of schools (Figure 22d) was concerned with the special needs of children. This was reflected in emphasis upon student counseling, cultural heritage, and intergroup relations programs. There were fewer disadvantaged students than in the other school groups. Classes tended to be small and non-traditional; reading and math were not stressed, and the instructional time of students in such activities was relatively low.

The outstanding characteristic of the fifth group of schools (Figure 22e) was the low level of district input into the decision-making process. Use was made

of nontraditional classrooms, with emphasis upon cultural heritage and interpersonal relations programs. In contrast with the previous group, class assignment was not based upon student needs.

The sixth group of schools (Figure 22f) had a relatively high level of disadvantaged students. There was school-wide program emphasis on reading and math, with considerable use of support personnel with small groups. This was accompanied by high levels of student exposure to math instruction.

As noted in the introduction to this subsection, one purpose in developing the new composite indices of school, program, and student characteristics was to determine what relationships these variables might have with student achievement measures. The results of analyses conducted to explore those relationships are discussed below in Section B4.

2. DERIVATION OF ALTERNATIVE OUTCOME (ACHIEVEMENT) MEASURES

One goal of the reanalysis activities was to derive new outcome measures, based on the achievement test data collected previously, that might show greater sensitivity to ESAA program effects than the raw and scaled test scores used in the earlier analyses. Although a number of different procedures for deriving the new measures were explored, the two techniques that appeared most promising involved (a) the selection of subsets of items that showed the greatest pre-to-posttest increase in percentages of correct responses, and (b) the calculation of revised summary or total scores using latent trait models.

Efforts to identify groups of items on which students showed particularly large gains were stimulated by a recent article by Carver (1974). In that article, it is argued that items for conventional standardized tests are selected more for their ability to discriminate among students with some common experiences than for their sensitivity to students' growth in an educational program. Carver contends that program evaluation can more meaningfully be accomplished by test items selected for their sensitivity to change. In the ESAA reanalyses, the only performance data available were from conventionally developed, standardized achievement tests, but it was felt that enhanced sensitivity to program effects might possibly result from selection of those standardized test items showing the greatest pre-post gain in percentages of correct responses for the treatment and control students combined. However, with one exception, the search of such items proved fruitless. For all but the mathematics computation subtest, the distribution of changes in item difficulties indicated all items about equally likely to change over the year. The exception to this was found in the mathematics computation subtest at level-3, the test used for the 4th and 5th grades. Here the items measuring performance in multiplication showed about a 7.5-point gain in percentages answering the item correctly between pre-and posttesting. This was in contrast to the three-to-four-point gain found in the other items' percentages. However, this identified set of multiplication items was not analized further as the items were found to be very difficult ones with a large apparent element of guessing in the responses.

A second approach involved the use of latent trait models to obtain more precise and possibly more sensitive measures of student achievement. These latent trait models assume that the likelihood of a person's getting an item correct is a function of his ability and of the item characteristic curve for that item. For the analyses done in this study, the item characteristic curve was defined by three parameters for each item: the probability of getting the item correct by chance, the item's difficulty, and its discriminating power. If the item characteristic curves are known for a set of items, then student abilities can be estimated from the set of responses to those items; these ability estimates based upon item characteristic curves are assumed to be superior to those obtained by more conventional test scoring procedures. Conventional scoring procedures use, as the basic estimate of ability or performance, a simple total count of the number of items answered correctly. In such a system all students who get 10 of 25 items correct are given the same score, regardless of which set of 10 items are answered correctly. By contrast, the ability estimates obtained through latent trait models are based upon the characteristics of the particular set of items correctly answered. Thus a student who answers 10 of the easier items correctly will get a lower score, or ability estimate, than the student who answers 10 more difficult items correctly.

In the past decade or so much work has been done toward finding practical techniques for obtaining both the ability estimates and the item characteristic curve parameters. The mathematical estimating procedure used in this study was developed by F. M. Lord (1974). The procedure uses the simple raw scores of the students as an initial estimate of abilities, and from the ability distributions obtains some initial estimates of the parameters of the item characteristic curves for each item. These estimated item parameters are then used to obtain improved estimates of the students' abilities, which are in turn used to obtain improved item parameters. Through such iterative procedures, stable estimates are sought of both the abilities and the item parameters.

Latent trait analyses were used for all of the elementary-level ESAA students for whom there were pre-and-posttests for the third year of the study. This required separate analyses for estimates of abilities for each subtest within

each test form, for both pre-and-posttesting for each of the three grade levels. The values of these 56 sets of estimates are not reported here because the procedures arbitrarily scale those ability estimates to an expected distribution with a mean of 0.0 and a standard deviation of 1.0: the distribution of values by themselves are not directly interpretable. With further standardizing, the scores were combined to yield mean Reading Total and mean Mathematics Total scores, pre and post, for each grade level. In order to reflect the gain, these final scores were standardized to the mean and standard deviations obtained on the ESAA scale scores for the same samples. The means and standard deviations of these latent trait scores and their correlations with the ESAA scale scores are shown in Table 64. The high correlations between the latent trait estimate and the ESAA scale score for each testing show that the new ability estimates were generally quite similar to the originally reported scores. However, it was felt that the differences in the two sets of scores, though small, might show meaningful effects, and the newly derived estimates were used in the reanalyses of relationships between program characteristics and outcome measures, as reported below in Section B4.

Table 64. Latent Trait Estimates and Correlations Between Those Estimates and ESAA Scale Scores, Based on Third-Year Test Data for Grades 3 Through 5

Grade 3

Latent-Trait Estimates

Variable	N*	Mean	Standard Deviation
L-T-READ-PRE	1289	511.68	102.80
L-T-MATH-PRE	1356	459.49	77.14
L-T-READ-POST	1297	550.13	102.47
L-T-MATH-POST	1300	524.58	91.36
ESAA-READ-PRE	1493	511.59	102.77
ESAA-MATH-PRE	1493	459.40	77.00
ESAA-READ-POST	1493	551.09	102.87
ESAA-MATH-POST	1493	524.60	91.67

Correlations

	Latent-Trait				ESAA-Scale			
	Read-Pre	Math-Pre	Read-Post	Math-Post	Read-Pre	Math-Pre	Read-Post	Math-Post
L-T-READ-PRE	1.00	0.72	0.88	0.66	0.95	0.65	0.87	0.64
L-T-MATH-PRE	0.72	1.00	0.68	0.75	0.60	0.88	0.72	0.73
L-T-READ-POST	0.88	0.68	1.00	0.69	0.80	0.63	0.96	0.63
L-T-MATH-POST	0.66	0.75	0.69	1.00	0.64	0.72	0.73	0.92
ESAA-READ-PRE	0.95	0.70	0.80	0.64	1.00	0.68	0.82	0.65
ESAA-MATH-PRE	0.65	0.88	0.63	0.72	0.68	1.00	0.70	0.77
ESAA-READ-POST	0.87	0.72	0.96	0.73	0.82	0.70	1.00	0.72
ESAA-MATH-POST	0.64	0.73	0.63	0.92	0.65	0.77	0.72	1.00

Table 64. Latent Trait Estimates and Correlations Between Those Estimates and ESAA Scale Scores, Based on Third-Year Test Data for Grades 3 Through 5 (Cont'd)

Grade 4

Latent-Trait Estimates

Variable	N*	Mean	Standard Deviation
L-T-READ-PRE	1274	567.95	103.59
L-T-MATH-PRE	1289	547.79	90.84
L-T-READ-POST	1302	599.01	99.24
L-T-MATH-POST	1316	588.34	101.77
ESAA-READ-PRE	1522	568.04	104.06
ESAA-MATH-PRE	1522	547.58	91.53
ESAA-READ-POST	1522	598.38	100.74
ESAA-MATH-POST	1522	589.16	102.09

Correlations

	Latent-Trait				ESAA-Scale			
	Read-Pre	Math-Pre	Read-Post	Math-Post	Read-Pre	Math-Pre	Read-Post	Math-Post
L-T-READ-PRE	1.00	0.69	0.78	0.60	0.90	0.66	0.77	0.62
L-T-MATH-PRE	0.69	1.00	0.66	0.71	0.68	0.87	0.69	0.73
L-T-READ-POST	0.78	0.66	1.00	0.68	0.78	0.66	0.88	0.66
L-T-MATH-POST	0.60	0.71	0.68	1.00	0.65	0.74	0.67	0.89
ESAA-READ-PRE	0.90	0.68	0.78	0.65	1.00	0.74	0.80	0.69
ESAA-MATH-PRE	0.66	0.87	0.66	0.74	0.74	1.00	0.72	0.81
ESAA-READ-POST	0.77	0.69	0.88	0.67	0.80	0.72	1.00	0.74
ESAA-MATH-POST	0.62	0.73	0.66	0.89	0.69	0.81	0.74	1.00

Table 64. Latent Trait Estimates and Correlations Between Those Estimates and ESAA Scale Scores, Based on Third-Year Test Data for Grades 3 Through 5 (Cont'd)

Grade 5

Latent-Trait Estimates

Variable	N*	Mean	Standard Deviation
L-T-READ-PRE	1242	625.66	101.32
L-T-MATH-PRE	1270	618.52	103.52
L-T-READ-POST	1257	645.38	100.93
L-T-MATH-POST	1272	651.07	107.38
ESAA-READ-PRE	1441	626.54	101.85
ESAA-MATH-PRE	1441	620.57	102.94
ESAA-READ-POST	1441	646.18	101.79
ESAA-MATH-POST	1441	651.52	107.79

Correlations

	Latent-Trait				ESAA-Scale			
	Read-Pre	Math-Pre	Read-Post	Math-Post	Read-Pre	Math-Pre	Read-Post	Math-Post
L-T-READ-PRE	1.00	0.69	0.85	0.66	0.94	0.69	0.85	0.66
L-T-MATH-PRE	0.69	1.00	0.67	0.77	0.71	0.92	0.70	0.78
L-T-READ-POST	0.85	0.67	1.00	0.69	0.83	0.67	0.94	0.69
L-T-MATH-POST	0.66	0.77	0.69	1.00	0.67	0.78	0.70	0.94
ESAA-READ-PRE	0.94	0.71	0.83	0.67	1.00	0.75	0.86	0.70
ESAA-MATH-PRE	0.69	0.92	0.67	0.78	0.75	1.00	0.73	0.83
ESAA-READ-POST	0.85	0.70	0.94	0.70	0.86	0.73	1.00	0.76
ESAA-MATH-POST	0.66	0.78	0.69	0.94	0.70	0.83	0.76	1.00

3. REANALYSIS OF DATA RELATED TO PROGRAM IMPACT

One of the major objectives of the ESAA evaluation was to study the ESAA program's impact upon student achievement. Extensive impact analyses were conducted on each study year's cross-sectional sample and upon the three-year longitudinal sample. A more extensive description of the outcome measures, the samples, and the results of these analyses is presented in earlier reports (see the list of documents at the end of this report).

A short summary of those results should provide a convenient framework on which to build the discussion of the impact reanalyses that were conducted. There was an indication of program impact on basic achievement (reading and mathematics) for the Basic Elementary sample but not for the Basic Secondary or the Pilot Elementary samples. Those results were consistent with the finding that only the Basic Elementary sample showed differences (favoring the treatment schools) in total student expenditures and resource allocations. The impact was evident only at grades 4 and 5. The grade 3 control group was initially at a significantly higher achievement level than the treatment group, making any interpretations of gains for that grade quite tenuous. The strongest evidence of impact was for the 4th-grade mathematics and 5th-grade reading, with all differences favoring the treatment group for these two grades.

The previous analyses used gains in ESAA scale scores to study the impact of the ESAA program upon student achievement. Covariate-adjusted ESAA scale posttest scores were also analyzed, giving results consistent with the gain score analyses. Those analyses used the school as the unit of analysis, that is, mean pretest and posttest scores for each school were used as the basic unit rather than scores for each individual student. The statistical method employed to test for treatment effects was analysis-of-variance.

As discussed in previous sections of this report, attempts were made during the contract extension period to construct new program indices and more refined outcome measures. Using these new indices and measures, program impact was reanalyzed to determine whether the new measures were more precise and whether any new information concerning program impact could be obtained. The reanalysis models, and results of those supplemental analyses, are discussed below.

a. Alternative Analysis Models and Outcome Measures

As indicated in a previous section, most of the methods used to seek new, more refined outcome measures were unproductive, but the ability scores derived from a latent trait model seemed to hold some possible promise as an alternative measure. Another possibility was to consider raw or unscaled scores, that is, the number of correct responses on the test. Since the formula used to construct scale scores is nonlinear and is built to try to equate different forms and test-levels, it was felt that possibly the standardization could have affected the variability of scores so that treatment and control differences were masked. Also, the presence of initial inequities in the achievement levels of the treatment and control groups suggested that the standardization transformation might have altered treatment and control differences. Thus, it was decided to use raw scores as well as latent trait-derived scores in the impact reanalyses.

None of the newly constructed program indices had large enough correlations with posttest measures, after controlling for pretest, or with gain to warrant an investigation of impact using these indices as covariates, or blocking variables for any of the new impact analyses.

Previous analyses used the school as the unit of analysis, as was appropriate from a strictly statistical viewpoint, since assignment to treatment and control conditions occurred at the school-level. While some of the reanalyses were conducted at the school-level to provide continuity with the analyses done in the past, other analyses were performed at the student-level. Although the student-level analyses may have violated certain basic independent assumptions relating to various statistical tests, such an approach seemed legitimate for exploratory purposes and for possibly revealing intra-school program effects that might have been obscured by strictly inter-school analyses.

Again, to provide continuity with previous analyses, gain scores were analyzed. In addition, analyses were conducted using posttest scores as the outcome measure. In analyses involving raw scores, raw gains would have been misleading, since the two forms of the test differed greatly in difficulty level. Thus, the raw score analyses used the test form as a covariate to control for differences in the distributions of scores from the two forms. This was not necessary

for analyses involving ESAA scale scores, or for the latent trait ability scores, since the algorithm used to construct these scores already took into account the form differences. Using posttest as the outcome, and controlling for pretest information, adds precision to the model when compared to gain score analyses. The difference between these two techniques can be illustrated as follows:

- Gain scores as outcomes

  $$POST - PRE = \mu + \alpha + \varepsilon$$

  where μ is the overall mean, α is the treatment effect, and ε is the residual or unexplained part of each score.

- Posttest scores as outcomes (controlling for pretest)

  $$POST - \beta * PRE = \mu + \alpha + \varepsilon$$

  where β is calculated to optimize the adjustment of the posttest score.

As can be seen in these two equations, these models will be similar if β is close to 1.0. Otherwise, interpretations based on the second model will probably be more accurate.

Previous analyses considered total reading and math scores separately. This involved conducting several (nonindependent) statistical tests and increased the possibility of spurious positive relationships. A technique which circumvents this problem and at the same time allows the test of more information is the multivariate analysis-of-variance (MANOVA). Conceptually it is similar to the analysis-of-variance, except that more than one outcome measure is considered. MANOVA addresses the question of whether there was an impact on achievement in general, rather than addressing specific subject areas. Specifically, it asks whether the treatment had an effect on any linear combination of the outcome variables. MANOVA takes the 'multiple test' nature of the procedure into account and adjusts the significance level accordingly. For these reasons, the MANOVA technique was also used in the reanalyses, with total reading and mathematics scores as outcome variables in one case, and reading and math subscale scores in another.

A widely recognized problem in impact analyses is the problem of interpreting impact when both treatment and control groups are changing over the evaluation

period, especially when the two groups are unequal at the start (see Bryk and Weisburg, 1977). Since this initial inequity was a potential problem in the ESAA Basic sample, additional analyses were conducted to investigate whether the apparent program impact was actually an artifact of the initial group inequities. In these analyses, an equation was constructed to predict gain using control group data, and that equation was used to calculate the treatment group's predicted gain. This latter predicted gain was then subtracted from the observed treatment group gain to yield an adjusted treatment effect. A second strategy involved finding the correlations of pretest and posttest with a dichotomous treatment-control (TC) variable (TC = 1 if treatment, 0 if control). A stronger correlation of TC with posttest would indicate a treatment effect. The correlation change from pretest to posttest was also tested to determine whether the effect was significant. This technique allows a test of impact which does not depend on the same assumptions as the traditional ANOVA, and hence provides a convenient, although conceptually more complicated, supplementary approach to these analyses.

b. Results of Reanalyses Using Third-Year Data

The Basic Elementary sample was the focus of the reanalyses, since earlier work had shown that this group was the only one where differential resource allocation had occurred between the treatment and control groups. In general, the reanalyses produced results similar to and consistent with those found in the earlier report covering the third evaluation year. Thus, the findings appear to support the interpretation that treatment-control differences were the result of program impact rather than being a chance artifact of these particular data. These reanalyses are discussed in more detail in the following paragraphs.

(1) Reanalyses Using ESAA Scale Scores

The previous analyses considered ESAA scale gain scores with the school as the unit of analysis. Grade 3 revealed that overall the control group gained slightly more than the treatment group, but the differences were not significant. At grades 4 and 5, achievement gains in reading and mathematics consistently favored the treatment group, and differences were significant for 4th-grade mathematics and 5th-grade reading (see Table 65). However, because of

initial inequities in the treatment and control groups, the possibility existed that these differences could be attributed to latent group growth differences rather than treatment. Several approaches to studying this issue were used.

First, the control group at each grade level was used to construct a linear growth curve (regression equation) to predict treatment group gain; this analysis produced the results in Table 65.

Table 65. Differences in and Observed Gains for the[2] Treatment Group Based on a Control Group Growth Curve

Grade	Observed Treatment Gain	Predicted Treatment Gain	T-C Raw Difference[1]	Adjusted Difference
Reading				
3	38.5	40.0	.2	-1.5
4	35.7	30.9	7.9	4.8
5	26.6	18.1	8.4*	8.5*
Mathematics				
3	61.2	66.9	-3.4	-5.7
4	46.9	37.9	9.9*	9.0*
5	36.8	30.5	6.0	6.3

1 These figures are from Table VI-1 of the Third-Year Report.
2 The school was the unit of analysis.
* A significant T-C difference ($p < .05$).

As can be seen in Table 65, use of the adjusted T-C differences does not change the interpretations of the results based on raw gains.

A second approach dealt with the correlations between pretests and a dichotomous variable (TC), indicating to which group the observation belonged (TC = 1 if treatment, 0 if control), and between posttest and TC. If the posttest-TC correlation was higher, it would indicate that high scores on posttest were more strongly related to the treatment group than had been the case for the pretest. Such a result could be interpreted as evidence of a treatment effect. The

correlation change was also tested to determine whether the effect was significant. Table 66 summarizes the results of this approach. Again the results are consistent with the previous analyses, and reinforce the interpretation of a positive program impact.

Table 66. School and Student Level Comparisons of Pretest-TC and Posttest-TC Correlations

Grade	Subject	School Level t^1	Student Level t^1
3	Reading	-.165	-.388
	Math	-.316	.398
4	Reading	.493	.863
	Math	1.843*	2.380*
5	Reading	1.410	1.755*
	Math	.222	.683

1 Hotelling's-t for testing the difference between two correlations.

* The correlation difference is significant ($\alpha < .05$).

The ESAA scale scores were also subjected to a multivariate analysis-of-variance (MANOVA), as described above. Models using total mathematics and reading scores were analyzed using both school and student as the unit of analysis. A student-level model was also constructed using the four subscale scores (two each in reading and mathematics) as the outcome measures. In each model, pretest scores were used as covariates in an attempt to control for initial differences and the impact of the treatment upon outcome measures was tested using the F-statistic, based on the Hotelling-Lawley trace. These F's and the corresponding probability levels are listed in Table 67.

The school-level analyses for the MANOVA do not reveal any significant treatment effect; however, at the student-level the same trends evident in the earlier analyses appear. Again, there appears to be no impact on the 3rd-grade achievement, but some evidence of impact at the 4th and 5th grades.

Table 67. ESAA Scale Score MANOVA Models

Grade	Total Scales School Level		Total Scales Student Level		Subscales Student Level	
	F	P	F	P	F	P
3	.14	.8612	1.25	.2869	.70	.5933
4	1.53	.2292	2.21	.1098	2.33	.0542
5	1.01	.3742	5.41	.0046	2.99	.0179

(2) Reanalyses Using Raw Scores

To determine whether the standardization algorithm used to calculate scaled scores influenced the results of the impact analyses, a MANOVA model was constructed using reading and mathematics total raw scores as outcome variables. Pretest information was again covaried out of posttest score (including pretest instrument level and form). This model was analyzed using the student as the unit of analysis. These results were generally consistent with those already reported, but added no new information.

(3) Reanalyses Using Latent Trait Ability Scores

As reported in a previous section, a latent trait ability score was calculated for each of the 5th-grade students on both the pretest and posttest. A MANOVA model was constructed using reading and mathematics posttest scores as outcome measures with pretest scores used as covariates. The resulting MANOVA F was 2.48 ($p = .084$). This is slightly less than the student-level MANOVA F's found using ESAA scale scores and raw scores, but still shows the same general trend of larger achievement gains for the treatment group.

In summary, the various impact reanalyses provided general support for the interpretations suggested by the third-year analyses. With the various approaches used to control for initial differences in treatment and control groups and to account for the multiple tests necessary for studying impact, the results still indicate that the ESAA program had a favorable impact upon the students in the program.

c. Results of Reanalyses Using Three-Year Longitudinal Data

A subset of the Basic Elementary sample was followed through three years of the ESAA evaluation. This subset had scores available at six points for both reading and mathematics. Its students were initially sampled in the first year of the ESAA evaluation and remained in the sample throughout the evaluation period, thus allowing the opportunity to study the impact of the ESAA program over an extended period.

In the initial year of the evaluation, matched schools were randomly assigned to treatment and control conditions. Within schools, students were selected who were deemed eligible to receive ESAA funds. This sampling procedure dictated use of the school as the unit of analysis since that was the unit on which random assignment occurred. However, because of attrition of both students and schools over the three-year evaluation period, the independence assumptions based on the initial sampling plan were no longer strictly satisfied. Moreover, the individual student, not the school, was the basic unit which was followed over this longitudinal period. To more directly address the issue of program impact on the students in the longitudinal sample, the individual student was chosen as the unit of analysis.

Figures 23 and 24 represent the profiles of this longitudinal sample of students for reading and mathematics. There is an important consistency in the two figures. During each school year, the treatment group gained more than the control, but between years the control gained more than the treatment. Since the between-years periods included summer vacation (non-treatment periods) this suggests that the 'latent' growth curve of the control group actually had a greater slope than the treatment group, and that in the absence of any treatment, the two groups would be diverging, i.e., the control group would be pulling further ahead.

The treatment group profiles in Figures 24 and 25 have the same general shape found by Pelavin & David (1977) in an analysis of the longitudinal data from several compensatory education programs. They conclude that inclusion of between-year periods within the longitudinal analyses tends to "considerably reduce

Figure 23. Longitudinal Profile - Reading

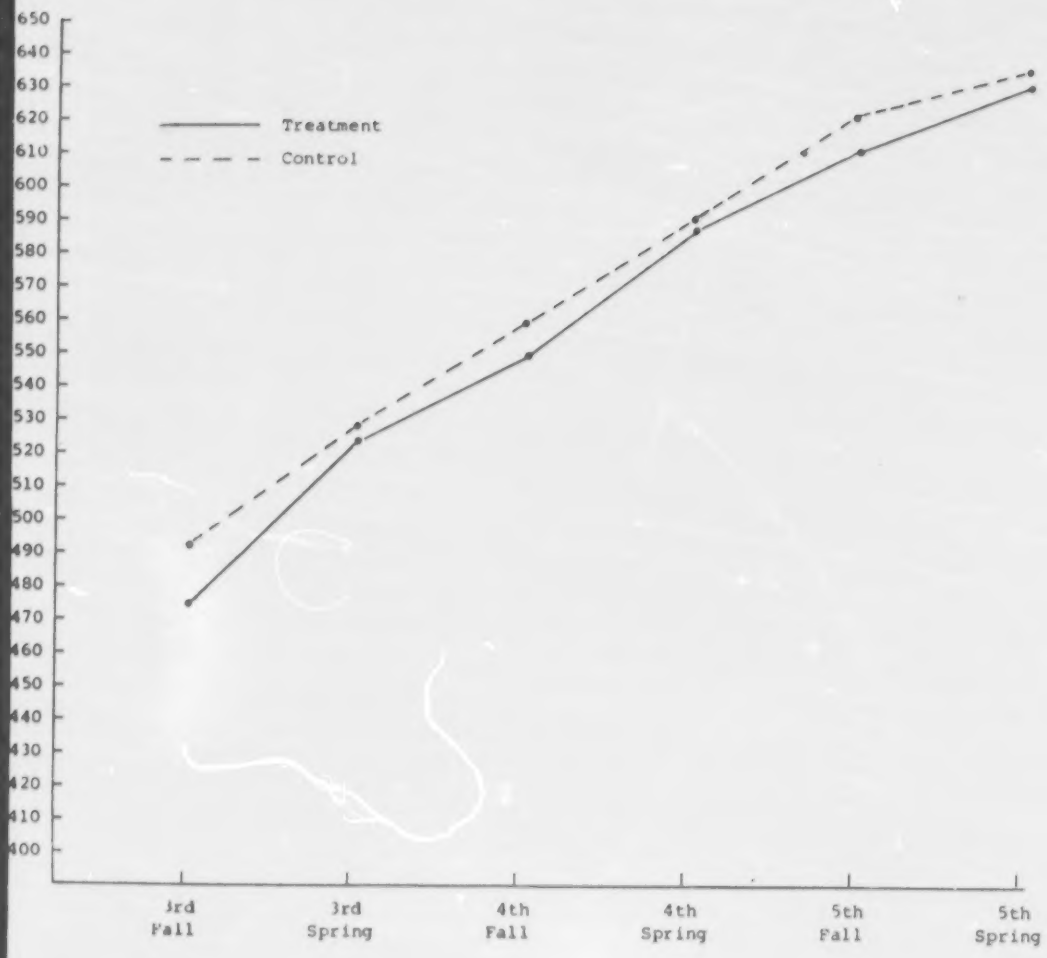
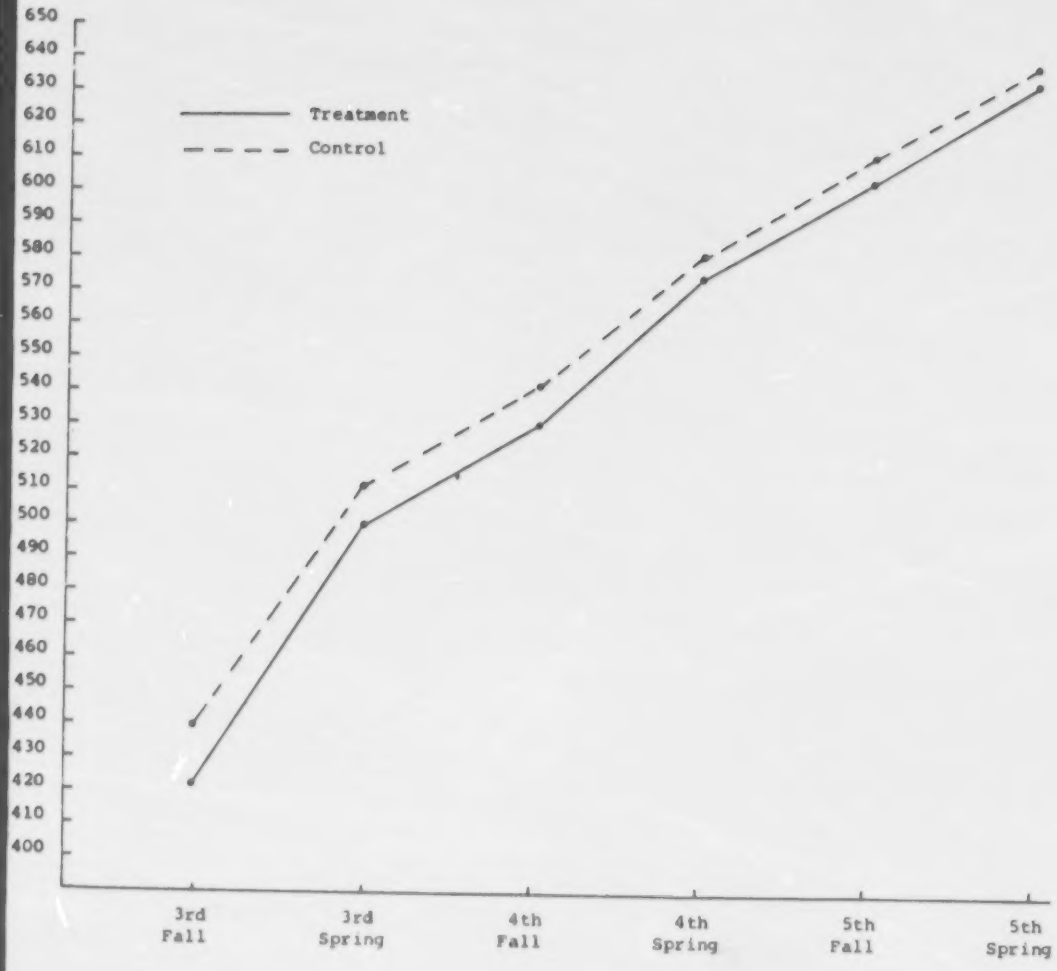


Figure 24. Longitudinal Profile - Math



estimates of achievement". As recommended by these authors, the ESAA longitudinal evaluation did include both fall and spring test scores over a period of more than one year. Thus, the ESAA data at least partially reflect the extent to which achievement is sustained.

The Sign Test can be used to test the hypothesis that the slope for the treatment is greater than for the control during the school year and that the slope for the treatment is less than for the control between school years. Of the ten slope comparisons that can be made, nine were in support of the hypothesis; in the tenth case the slopes were identical. In the population, if there were really no differences in the slopes at the various parts of these profiles, the probability of this pattern occurring would be less than .01 (sign test). Thus, the results strongly suggest a positive treatment effect in the longitudinal sample.

A second method of studying program impact for the longitudinal sample is to compare treatment and control gains over the entire three-year period. Table 68 summarizes these comparisons, again using the individual student as the unit of analysis.

Table 68. T-C Comparisons for the Longitudinal Sample Over the Entire Three Years of the Evaluation (ESAA Scale Scores)

	Treatment Gain(T)	Control Gain(C)	T-C	t[1]
Mathematics	213	201	12	2.20*
Reading	157	143	14	2.86*

1 Student's t-test for differences in means
* The difference is significant ($p < .05$).

These results add strong support to the hypothesis of a positive ESAA program impact on student achievement. Although the effects appear to be small for any given year, the gains consistently favor the treatment group. There also appears to be a cumulative effect over the three-year period. It is a question of interpretation and judgment whether these gain differentials are 'educationally

significant'. They do represent about seven percent to ten percent of the total gain and, if the pattern were to remain consistent throughout a student's entire school exposure, the overall effect would be impressive.

d. Summary

For the three-year cross-sectional sample, the use of alternative analysis models and the search for more refined outcome measures did not produce any dramatic new results. Rather, they tended to reinforce the finding of a positive ESAA program effect at the 4th and 5th grades suggested by the earlier analyses of the data.

The reanalyses of the Basic Elementary longitudinal sample produced results which much more strongly supported the interpretations of a favorable ESAA program impact. These reanalyses used the individual student as the analysis unit and directly addressed the question of whether the students in the ESAA program for three years gained more in reading and mathematics than the set of students in the control group who had no opportunity to benefit from the ESAA program. (It should be noted that this is different than asking whether the average achievement scores in the treatment schools increased more than the average achievement scores in the control schools.) The student-level analyses indicate that there was a significant program effect in both reading and mathematics in the longitudinal sample.

In summary, there does appear to be evidence of a positive ESAA program impact in both the 4th-and-5th grade cross-sectional samples. The longitudinal sample, when analyzed at the student-level, offers strong evidence of program impact over a three-year period.

4. REANALYSES OF DATA ON FACTORS INFLUENCING ACHIEVEMENT GAINS (RELATIONAL ANALYSES)

The relational models discussed here used the newly derived outcome and predictor measures (described in Sections B1 and B2) in a set of analyses paralleling some of those conducted previously for each of the three years' evaluations. It was hoped that the new analyses would substantiate the earlier findings, and possibly clarify some of the relationships between program factors and student outcomes. Several preliminary analyses were performed using various groups of the new predictor measures; the present discussion focuses on the subset of analyses that yielded the clearest set of relationships.

Tables 69 and 70 show the statistics from a set of three regression analyses on reading outcomes, and three analyses of the mathematics achievement outcomes. The data for Table 69 are based on the third-year Basic Elementary sample students for whom there were data linking each student to a regular reading teacher; in other words, this is the third-year sample of elementary Basic Program students for whom there were achievement gain measures and classroom-level predictor variables related to their reading programs. Similarly, the sample for Table 70 consists of elementary Basic Program students in the third-year study for whom there were pretest and posttest mathematics scores and teacher-associated predictor variables. The tables show the significant regression coefficients obtained when the achievement posttest measures were predicted from a set of variables that included the achievement pretest measures, the student and teacher attitude measures, the three exposure measures, and the four second-order program indices reflecting instructional emphasis. The outcome measures used in the regressions are the latent trait estimates of the reading and mathematics achievement at posttesting. The latent trait estimates of the reading and mathematics pretest achievement are used as predictors in these models. (The derivation of the ability estimates of the achievement scores is discussed above in Section B2, while the new predictor measures are described in Section B1.) Stepwise regression was used for the analyses.

As expected, both tables show that the prediction equations are dominated by the pretest measures, which account for 95 percent of more of the total explained

variance. Of the remaining predictors, the clearest pattern is seen in the relationship of the student attitude index, STATD, to the posttest outcome. A significant positive association is seen at the 3rd and 4th grades in reading (Table 69), and at all three elementary grades in mathematics (Table 70). These results are consistent with those found in the prior analyses. This composite measure of student attitude includes the variables reflecting the student's liking for school, as well as the locus of control variables positively associated with outcomes in the prior analyses.

A weaker, but perhaps more important, association in these results may be seen in the pattern of relationships of the mathematics exposure measure, TOTMATH, to the outcomes. The mathematics exposure index is an adjusted measure of the number of minutes spent in mathematics instruction. Small but significant positive associations are seen between the exposure to mathematics instruction and the mathematics achievement measure at posttest, independent of the pretest level in mathematics. On the other hand, Table 69 shows that the associations between exposure to mathematics instruction and reading outcomes are negative. These associations are found for the 3rd and 4th grades only, reflecting again the pattern found in the third year's evaluation of very low predictive power in the 5th grade.

The pattern of associations seen in these additional analyses is similar to, but clearer than, patterns found in the second year and third year of the evaluation. These new data would appear to substantiate the earlier interpretation that students who have more math instruction show greater gains in mathematics achievement. The negative associations between the exposure to mathematics instruction and the gain in reading achievement are difficult to explain, however, since an increased exposure to mathematics did not necessarily mean decreased exposure to reading instruction. Indeed, the data show positive correlations of 0.32 to 0.42 between mathematics and reading exposure in these samples.

Overall, these new relational analyses are highly consistent with results found in the earlier studies, but give stronger evidence of a positive relationship between mathematics exposure and mathematics achievement. The strenghtening

of this relationship may have resulted in part from the greater precision of the new exposure indices, as well as from the use of alternative mathematics outcome measures (latent trait estimates) having slightly smaller pre-to-post correlations than did the earlier scale scores. These lessened correlations may indicate a small but meaningful increase in the sensitivity of the new measures to growth. It seems altogether possible that, had even more sensitive and more precise predictors and outcomes been available, they would have shown additional program variables having effects on student outcomes.

Table 69. Significant Regression Coefficients in Models Predicting Reading Achievement, Posttest Measures From Selected Predictors, and Pretest for Grades 3, 4, and 5

		Grade 3	Grade 4	Grade 5
R^2 (with all predictors)		0.762	0.630	0.731
R^2 (with pretest only)		0.758	0.619	0.728
N		1075	965	876
Intercept		126.282	161.509	131.617
Predictors				
ABL1RTOT	(Estimate reading achievement, pretest)	0.869	0.758	0.852
STATD	(Student attitude composite)	0.773	2.263	---
TOTMATH	(Mathematics exposure index)	-0.001	-0.264	---
TCHATD	(Teacher attitude composite)	-1.204	-0.407	---
TOTSCH	(School exposure index, all instruction)	---		-0.031
SUBJREAD	(Index of reading subject matter emphasis)	---		-0.446

Table 70. Significant Regression Coefficients in Model Predicting Mathematics Achievement Posttest Measures from Selected Predictors and Pretest for Grades 3, 4, and 5

		Grade 3	Grade 4	Grade 5
R^2 (with all predictors)		0.574	0.524	0.593
R^2 (with pretest only)		0.561	0.498	0.579
N		1167	1051	913
Intercept		167.035	298.612	192.441
Predictors				
ABLIMTOT	(Estimate of mathematics achievement, pretest)	0.859	0.788	0.778
STATD	(Student attitude composite)	1.159	1.999	1.599
TOTMATH	(Mathematic exposure index)	0.001	0.002	---
TCHATD	(Teacher attitude composite)	-1.664	-1.750	---
READGRP	(Emphasis on small reading groups)	0.463	-0.649	---
TOTSCH	(School exposure index, all instruction)	---	-0.059	---
SUBJREAD	(Index of reading subject matter emphasis)	---	0.429	---

IV. ADDITIONAL ANALYSES OF IN-DEPTH STUDY DATA

This section describes the results of additional analyses performed on data from the in-depth studies conducted in the second and third years of the overall ESAA evaluation (1974-75 and 1975-76). Those studies were conducted on subsets of schools from the total ESAA evaluation sample selected for maximal variation in their degree of success in raising student achievement. Observations of school and classroom activities and interviews with teachers and principals were carried out during two-week visits to each in-depth sample school by specially trained data collectors/observers. In addition, questionnaires were administered to students and staff members. There were 24 elementary schools in the 1974-75 in-depth sample, and 26 elementary schools in the 1975-76 sample.

The first set of supplemental analyses discussed below was conducted in an effort to clarify and further explicate some of the findings reported earlier for the in-depth studies (Wellisch et al, 1976; Wellisch et al, 1977), concerning the apparent importance of school organization and management in determining school success. In particular, we report on findings concerning the effects of different levels of school coordination in moderating the strength of relationships between school resource utilization and school success.

The second subsection describes efforts to define causal models that might link school variables to student achievement via classroom events. Particular emphasis is placed here on attempts to establish links between certain classroom events and the school-level organizational and management variables that had previously been found to associate with variations in school success. Because of the importance of the classroom observational data in these analyses, the discussion of causal models is preceded by a discussion of activities performed to identify and compensate for possible systematic observer bias.

A final set of reanalyses used data from both in-depth studies. The design of these studies made it possible to compare results using different definitions of school success (based on prior achievement level, prior gain and study year gain), and so to examine more critically the direction of a possible causal relationship between teachers' satisfaction and student achievement.

A. RELATIONSHIPS AMONG RESOURCE USE, PROGRAM COORDINATION, AND SCHOOL SUCCESS

In voting to increase funds for schools, citizens and legislators have assumed that more and better resources would, in general, improve the quality of education. But in view of the variety of resources that are available and the diversity of circumstances in which schools operate, it has also been recognized that the relationship is probably not so simple as "More resources result in better schools".

The 1975-76 ESAA In-Depth Study found that use-costs for several categories of resources were higher in successful schools than in nonsuccessful schools, but few of these differences were statistically significant. One limitation of these original analyses was that none of them took into account contextual differences, such as how the schools were organized and managed. This report presents the results of an effort to extend the inquiry along such lines by reanalyzing the 1975-76 ESAA in-depth resource use data, while addressing the following question: Does school organization and management condition the impact of resource use on school success in raising student achievement?

Two organizational factors were examined: administrative leadership in instruction, and the extent to which instructional programs were coordinated throughout the school. As reported in the 1975-76 ESAA In-Depth Study, (Wellisch et al, 1977), both factors were directly related to school success. However, in the reanalysis only program coordination was found to condition the impact of resource use on school success in a consistent manner. Therefore, the present report focuses exclusively on findings concerned with the conditioning impact of the extent to which instructional programs were coordinated throughout the school.

The report is organized in three main sections. The first section reviews relevant material from the 1975-76 ESAA In-Depth Study. Next is a methodology section that outlines the types of analyses used to explore the relationships among program coordination, resource use, and student achievement. A third section presents the results. Following these three main sections is one that summarizes and briefly discusses the results.

1. OVERVIEW OF THE 1975-76 IN-DEPTH STUDY

The 1975-76 ESAA In-Depth Study sample consisted of 26 schools selected from 78 elementary schools that were participating in the third and final year of the national evaluation of the Emergency School Aid Act (ESAA). The purpose of the in-depth study was to identify school and classroom determinants of success in raising academic achievement among students who were largely minority and disadvantaged. The following paragraphs review relevant material from the study: data sources and collection procedures; the criteria used to define school success; the descriptive and analytic results of program coordination; and resource use methodology.

a. Data Sources and Collection

Each of the 26 in-depth study schools was visited by a two-person interviewer/observer team for two weeks during the winter of 1976. During that time, the teams administered questionnaires and interviews to principals and teachers, and used structured protocols to record observations. The analysis of resource use relied primarily on questionnaire data and, to a lesser extent, observational data. Some resource data were also obtained from district offices by mail and by phone. Assessments of program coordination were based on the interviews with teachers and principals. The California Achievement Tests (CAT) (levels 2 and 3, Forms A and B) were administered in the early fall and late spring of 1975-76 to measure achievement.

b. School Success in Raising Student Achievement

School success in raising student achievement was defined in terms of the number of grades that gained in national percentile standing between the fall of 1975 and spring of 1976. ESAA sample students in three grades (3, 4, and 5) were tested in two subject areas, reading and math. Schools were classified successful if (1), at least two grades gained in national percentile standing in one of the two subject areas and, (2), at least one grade gained in the other subject

area.* Schools were classified nonsuccessful if no more than one grade improved in either area. These criteria were designed to produce well-separated comparison groups, that is, to exclude schools that were neither clearly successful nor clearly nonsuccessful. Application of these criteria identified nine successful and thirteen nonsuccessful schools. Of the remaining four schools, three met neither the success nor nonsuccess criteria. The fourth school was excluded because data on only one grade were available.

For reanalyses using multiple regression techniques, a continuous rather than dichotomous outcome measure was appropriate. As a result, a measure of school success based on raw gains, standardized by grade and on the sample, was also used. This measure represented the school average gain of the sample students in the two subject areas, reading and math. Since classification procedures were not required in computing school average gains, the three schools that did not meet the success/nonsuccess criteria were not excluded from analyses using the continuous measure of school success. Consequently, results that refer to school average gain are based on a sample of 25 schools, whereas results that refer to school success, or successful and nonsuccessful schools, are based on 22 schools.

c. Program Coordination

Teachers and principals were asked to what extent their school's instructional programs were coordinated. Interviewers chose one of three options to summarize each individual's response: to a large extent, to some extent, or very little. 'A large extent' referred to programs that were coordinated in many aspects--content, sequence of objectives, and use of materials--and throughout all grades. 'Very little' described schools in which each teacher generally planned an instructional program for his or her own class. The principal's and teachers' responses for a given school were combined, resulting in scores having a possi-

*In five schools, gain scores were available for only two grades. Classification of three of these schools was unambiguous since all grades either gained or dropped in both subject areas. Classification of the other two was judgmental (see p. II-7 in An In-Depth Study of Emergency School Aid Act (ESAA) Schools: 1975-1976, by Wellisch et al., which is available from Dr. David Schoemaker, USOE).

ble range of 0 to 4, with 4 representing maximum coordination. The distribution of these scores is shown in Table 71.

Table 71. Distribution of Scores Measuring the Extent to Which Instructional Programs Were Coordinated Throughout the School

Interpretation of Scores	Scores Median = 3.1	Number of Schools * (N = 25)
Coordinated To a Large Extent	4	4
.	3 to 3.9	10
.	2 to 2.9	9
.	1 to 1.9	2
Coordinated Very Little	0 to 0.9	0

*The coordination scores of the three schools that were excluded from analyses using the 'school success' outcome (but not excluded from analyses using 'school average gain') were 1.0, 2.5, and 4.0.

Based on these staff estimates, successful schools were significantly more likely to report highly coordinated instructional programs ($\alpha \leq .01$, Mann Whitney U test, N = 22). The relationship between the extent of program coordination and school average gain was positive and substantial--the greater the coordination, the greater the gain: ($r = .53$, $\alpha \leq .001$, N = 25).

d. Resource Use Methodology

The dollar figures reported in this document are estimates of actual use-costs, or allocations of resources within the school and classroom. This approach to measuring school resources is very different from one that uses expenditure data. The expenditure approach relies primarily on financial data from school district accounting records. The resource allocation approach, on the other hand, examines the kinds and amounts of resources in use in the school and classrooms.*

*A detailed account of the methodology can be found in Appendix D of An In-Depth Study of Emergency School Aid Act (ESAA) Schools: 1975-76 by Wellisch et al., available from Dr. David Shoemaker, USOE.

For a general discussion of the resource allocation approach to program cost analysis, see S. A. Haggart, Program Cost Analysis in Educational Planning. The Rand Corporation, P-4744, 1971.

Different resources required different procedures to convert use data into cost estimates. In general, an average cost per unit was assigned to each resource, and a total use-cost figure for the school was obtained by multiplying the per-unit cost by the number of units, or amount of the resources in use. Per-pupil or per-teacher costs were calculated where appropriate. The resources with which the in-depth study was primarily concerned, and the procedures used to obtain the dollar estimates associated with the use of each were:

- Teaching Staff - Costs for the staff who provided reading and math instruction--regular teachers, reading and math specialists, and instructional aides--were estimated by multiplying the proportion of time spent instructing these subjects by an average annual salary obtained from district questionnaires for each type of staff.*
- Support Staff - Costs for staff providing support services to reading and math teachers--principals, counselors and psychologists, speech therapists, and others--were estimated by multiplying the amount of service teachers reported receiving by an average hourly wage for each type of support staff.*
- Equipment - Costs for equipment such as tape recorders, record players, projectors, and controlled readers were calculated on the basis of an average annualized capital and operating cost for each resource. These cost factors were multiplied by the proportion of time each teacher used the resource in reading and math.
- Staff Development - Inservice training costs for teachers and principals were estimated by multiplying their respective average hourly rates by the total hours of training they received during the year.*
- Program Development - Costs for the time teachers and principals devoted to assessing needs and to developing and evaluating instructional programs were estimated by multiplying their respective average hourly rates by the total hours they reported spending in these activities during the year.*

---

*This is an average of district averages.

2. METHODOLOGY FOR REANALYSES

The same categories of resources that were reported in the 1975-76 ESAA In-Depth Study were reanalyzed. However, instead of two separate analyses, one for reading and one for math, a single analysis that combined the two basic skill areas was performed. One reason for this approach was that no available information on program coordination made any distinction between the coordination of reading programs and the coordination of math programs. Secondly, a substantial part of the analysis required comparing subgroups of schools. As will be seen, the small size of some of the subgroups hampered our ability to interpret results; separate reading and math analyses would have made the problem even more severe.

Cost figures for various resources were broken down by level of coordination and school success, that is, schools were divided into two groups, those in which program coordination was low and those in which it was high, and then the mean resource use-costs for successful versus nonsuccessful schools were compared in each of the two groups. 'Low' program coordination was defined by scores below the median, and 'high' program coordination by median scores and above. Table 72 shows the number of schools in each of the four subgroups: (1) low coordination/nonsuccessful, (2) low coordination/successful, (3) high coordination/nonsuccessful, and (4) high coordination/successful.

Table 72. Distribution of Schools by Level of Program Coordination and School Success

		School Success		
		Nonsuccessful	Successful	Total
Program Coordination	Low	8	2	10
	High	5	7	12
	Total	13	9	22

In addition, multiple regression techniques were used to evaluate the relative contributions of program coordination and resource use-costs in predicting school average gain; to assess the interaction between the two variables; and to evaluate the predictive capacity of their joint effects. Multiple regression also provided estimates of the proportion of variance in school average gain that was explained by these variables.

3. RESULTS

a. Comparisons of Average Resource-Use Costs in Different Subgroups of Schools

Comparisons of resource use in school with highly coordinated programs and those with less coordinated programs yielded two general findings:

- For certain categories of resources, the strength of relationships between resource use and school success was different for highly coordinated schools and less coordinated schools. In most cases, there was a stronger relationship in schools with highly coordinated programs, but for a few resources the relationship was stronger in schools with less coordinated programs.
- For certain other resources, coordination level did not appear to affect the strength of the relationships between resource use and school success, but absolute cost levels for resources were different in the highly coordinated and less coordinated programs. In these cases, because costs were consistently higher for one coordination-level group than for the other, the relationships between level of resource use and school success were apparent only when tests of those relationships were performed separately for highly coordinated and less coordinated schools.

The remainder of this section describes in greater detail the study's findings regarding the two general categories of resources distinguished above, i.e., resources for which the strength of relationships between resource use and school success was different for highly coordinated and less coordinated schools, and those for which absolute cost levels for resources were different in the two coordination level groups.

(1) Resources for Which Success/Resource-Use Relationships Depended on Coordination Level

(a) Relationships Stronger in Schools With Highly Coordinated Programs. For one set of resources, ones associated with school principals, the relationships between use-costs and school success was stronger in the highly coordinated schools. Specifically, this was true for resource-use-costs involving:

- the time the principal devoted to program planning, needs assessment, and program evaluation;
- the time the principal devoted to curriculum development; and
- the amount of inservice training the principal received in how to develop and manage instructional programs.

Average use-costs for these principal-related resources in different subgroups of schools may be compared by inspecting Tables 73 through 75. In these tables, the average cost of each subgroup of schools is based on the number of schools shown in the corresponding position in Table 72. For example, it is evident that for the sample as a whole, principal costs for program planning, curriculum development, and inservice training were somewhat higher in successful schools, but differed little between schools with highly coordinated programs and schools with less coordinated programs. However, when the average costs of successful versus nonsuccessful schools were compared separately for the two groups of schools--highly coordinated and less coordinated--the relationship of these resource costs of success was substantially stronger in the highly coordinated groups. Since these principal-related resources were strongly intercorrelated, a single cost for the principal's time in these activities was obtained by summing the three (see Table 86). Comparisons of the average overall costs of the principal's time in developing and managing instructional programs and acquiring related skills are shown in Table 76. Within the subgroup of schools where programs were highly coordinated, the difference between successful and nonsuccessful schools in these overall principal costs was substantial and significant; they averaged well over twice as much in successful schools, $2,195, as in nonsuccessful schools, $831.

Table 73. Average Cost of the Principal's Time in Program Planning, Needs Assessment and Program Evaluation

		School Success			
		Nonsuccessful		Successful	
Program Coordination	Low	$576		$576	$576
	High	$323	*	$735	$563
		$479	*	$700	$569

Table 74. Average Cost of the Principal's Time in Curriculum Development

		School Success			
		Nonsuccessful		Successful	
Program Coordination	Low	$577		$653	$592
	High	$231	*	$735	$525
		$444		$717	$551

*Difference between the two means on either side of the asterisk is significant at $\alpha \leq .05$, with a one-tailed t-test.

Table 75. Average Cost of the Principal's Time in Inservice Training

		School Success		
		Nonsuccessful	Successful	
Program Coordination	Low	$577	$807	$623
	High	$277	$725	$538
		$462	$743	$572

Table 76. Average Overall Cost of the Time the Principal Spent in Instructional Management

		School Success			
		Nonsuccessful		Successful	
Program Coordinaiton	Low	$1730		$2036	$1791
	High	$ 831	*	$2195	$1627
		$1384		$2160	$1702

*Difference between the two means on either side of the asterisk is significant at $\alpha \leq .05$, with a one-tailed t-test.

Successful schools with extensive program coordination were characterized not only by the high use of the principal's time in program planning, but also by the low use of teachers' time in this activity. In the sample as a whole the two resources were inversely related: the greater the principal's time in program planning, the less were teachers involved in planning programs ($r = -.35$, $\alpha \leq .05$, $N=25$). Among highly coordinated schools, the relationship was even stronger ($r = -.53$, $\alpha \leq .03$, $N=13$). As may be seen in Table 77, teachers spent the least time in program planning in successful schools with highly coordinated programs.

Table 77. Average Per-Teacher Cost of the Teachers' Time in Program Planning

		School Success			
		Nonsuccessful		Successful	
Program Coordination	Low	$14.00		$15.89	$14.38
	High	$15.39	*	$11.03	$12.87
		$14.53		$12.11	$13.55

*Difference between the two means on either side of the asterisk is significant at $\alpha \leq .05$, with a one-tailed t-test.

(b) Relationships Stronger in Schools With Less Coordinated Programs. Successful schools where program coordination was not extensive and teachers planned instruction for their own classes more independently of one another were characterized by greater resource allocations to equipment, to instructional aides and to instructional time of regular teachers (Tables 78, 79, and 80). However, this description is very tentative since only two of the ten schools of this type were successful and differences were statistically significant for only one of these resources: the teacher's time in instruction.*

Per-pupil costs for the time teachers spent instructing reading and math were greater in successful than in nonsuccessful schools, in the sample as a whole and in the two subgroups, but especially in the subgroup where program coordination was less extensive and teachers planned instruction for their own classes. The difference, an average of $144 in successful schools versus $110 in nonsuccessful ones, was statistically significant (see Table 78).

Table 78. Average Per-Pupil Cost of Teachers' Time in Reading and Math Instruction

		School Success			
		Nonsuccessful		Successful	
Program Coordination	Low	$110	*	$144	$117
	High	$105		$114	$110
		$108		$120	$113

*Difference between the two means on either side of the asterisk is significant at $\alpha \leq .05$, with a one-tailed t-test.

*Because of the small sample size, a nonparametric 'equivalent' of the t-test, the Mann Whitney U test, was used to assess the significance of the differences between the successful and nonsuccessful groups where program coordination was low. In a sample as small and as skewed as this (8 nonsuccessful and 2 successful) differences were significant only when both successful schools ranked at one extreme of the distribution.

Relationships were stronger but not statistically significant in the subgroup of groups with less coordinated programs for two other resources: equipment and instructional aides. The differences are shown in Tables 79 and 80.

Table 79. Average Per-Pupil Cost for Equipment Associated With Reading and Math Instruction

		School Success		
		Nonsuccessful	Successful	
Program Coordination	Low	$ .91	$1.31	$ .99
	High	$ .60	$ .66	$ .64
		$ .80	$ .80	$ .80

Table 80. Average Per-Pupil Cost of Instructional Aides Assisting Teachers With Primary Responsibility in Reading and Math Instruction

		School Success		
		Nonsuccessful	Successful	
Program Coordination	Low	$42	$86	$51
	High	$44	$28	$35
		$43	$41	$42

While larger allocations of the principal's time to instruction-related activities may be effective where programs are coordinated schoolwide, larger allocations to equipment, aides, and teacher's time may be particularly appropriate in schools where programs are not extensively coordinated. Unlike resources associated with the principal, these three resources should be amenable to the efficient use and management by individual teachers. For example, large investments in equipment may show returns in schools where teachers plan instruction independently of one another--sufficiently large, that is, such that use

is not hampered by scheduling conflicts. On the other hand, schools may also invest in scheduling itself. Since procedures and personnel to handle scheduling may pre-exist as a result of program coordination, equipment may not be so critical a resource in those schools where coordination is emphasized.

(2) Resources Whose Use Differed for Highly Coordinated and Less Coordinated Schools

For the resource variables discussed above, the degree of program coordination apparently affected the strength of relationship between resource-use costs and school success. For certain other resource variables, concerned with teacher inservice training, the degree of program coordination was not a factor in determining relationships between resources and school success, but was associated with differences in the cost-level or manner of resource use. Specifically, as shown below in Table 81, there was a lower average per-teacher expenditure for inservice training in highly coordinated schools ($5.52) than in less-coordinated schools ($7.63), and this difference was significant at the .05 level.

Table 81. Average Cost Per Teacher for Inservice Training in Reading, Math, and in Teaching Disadvantaged Students[a]

		School Success		
		Nonsuccessful	Successful	
Program Coordination	Low	$7.46	$8.31	$7.63 *
	High	$4.85	$6.07	$5.52
		$6.45	$6.63	$6.52

[a]Total N of schools = 21. High Coordination/Successful Group N = 6.

*Difference between the two means on either side of the asterisk is significant at $\alpha \leq .05$, with a one-tailed t-test.

Teacher inservice training was also examined in terms of the time various personnel spent providing that training.* Table 82 lists these personnel and shows the average time of each in schools with highly coordinated programs compared with their time in schools with less coordinated programs.

Table 82. Average Hours Various Personnel Delivered Teacher Inservice Training by Type of School[a]

	Type of School	
Personnel	High Program Coordination	Low Program Coordination
Principal	52 hours	20 hours
Project Director	8 hours	13 hours
Teachers	22 hours	11 hours
Consultants	10 hours	21 hours
Guest Speakers	3 hours	8 hours
Publisher Representatives	3 hours	6 hours
District Staff	11 hours	52 hours
Total Delivery Hours	109 hours	130 hours

[a]Data from Principal Questionnaires

Total amount of inservice training time provided by the personnel was not significantly related to school success or to the degree of school program coordination. Furthermore, it was not related to the amount of inservice training that teachers reported themselves to have received; this would seem to indicate that schools varied in the intensity of delivery (ratio of inservice instructors to teacher trainees). However, statistically significant relationships were found when the different sources of inservice training time were separately examined. Table 83 shows the relative percentages of teacher inservice training provided by on-site personnel (principals and teachers who were part of the

*Since salary schedules for some of these personnel were not available, the cost of this resource was not estimated.

schools concerned). It can be seen that 66 percent of the training in highly coordinated schools was provided by on-site personnel, compared with only 35 percent for less coordinated schools: this difference is significant at the .01 level. In addition, it may be noted that more on-site personnel were used in successful schools (64 percent) than nonsuccessful schools (43 percent): this difference is significant at the .05 level.

Table 83. Average Percentage of Inservice Delivery Time Attributed to On-Site Personnel (the Principal and Teachers)

		School Success		
		Nonsuccessful	Successful	
Program Coordination	Low	32%	47%	35%
				**
	High	61%	70%	66%
		43%	* 64%	

*Difference is significant at $\alpha \leq .05$, one-tailed t-test.

**Difference is significant at $\alpha \leq .01$, one-tailed t-test.

b. Models of School Average Gains in Student Achievement

Several models of school average gains were examined in an effort to identify the combination of program and resource variables best explaining school average gains.* In each model, gain was posited as a function of three variables: program coordination, the use-cost of a particular resource, and the interaction between these two. Being relatively simple, the models will be referenced by the resource in question, each of which has been discussed and tabled earlier in this section.

*The term "model" refers to a mathematical description that is a linear function of the form, $Y = A_0 + A_1X_1 + A_2X_2 + \ldots + A_nX_n$, where Y is the estimated gain, the A_i's are constants estimated by a least squares procedure and the X_0's are predictor (independent) variables.

The models were evaluated in several ways: (1) by the explanatory power of the full model, in which all three variables were used to predict school average gain; (2) by the unique contribution of each variable, the interaction variable being of particular interest; and (3) by the explanatory power of models restricted to fewer than three variables (when the contribution of the interaction variable was not statistically significant).

In general, the variance in school average gain in student achievement that was explained by these models ranged from 25 percent to 46 percent. Full models in which more than 30 percent of the variance was accounted for were statistically significant. These included the models in which the resources were equipment (see Table 78), models involving the teachers' time in reading and math instruction (see Table 80), and models involving resources of the type found more commonly in successful schools having extensively coordinated programs (see Tables 73 through 77). In the remainder of this section, only the 'best' model--the one explaining the largest amount of outcome variance--will be discussed. Summary statistics on the other models are shown in Tables 92 and 94 appended.

In the 'best' model, summarized below in Table 84, 46 percent of the variance in school average gain in student achievement was accounted for by a model using the following three predictors:

- The extent to which the instructional program was coordinated throughout the school;
- The overall time the principal spent in instructional management (planning and evaluating programs, developing curriculum, and receiving in-service training in related skills); and
- A variable representing the unique, additional effect of the combination of these two, that is, an interaction variable.

It can be seen in Table 84 that each of the two component program variables (program coordination and principal's time) makes a statistically significant contribution to the outcome variance, as does the full model, including those

variables plus the interaction term. However, the interaction term itself does not contribute significantly to the variance in achievement gains. This latter finding is somewhat surprising, as other analyses discussed earlier have suggested meaningful interaction effects, e.g., the difference in strength of relationships between resource use and school success, for highly coordinated and less coordinated schools. Such effects seem also to be indicated by the correlations shown in Table 85.

Table 84. Analysis-of-Variance of School Average Gain in Student Achievement

Source of Variation	Proportion of Variance Explained	df	F Value
a. Program Coordination, Adjusted for Principal's Time	.29	1/21	11.04**
b. Principal's Time, Adjusted for Program Coordination	.12	1/21	4.81*
c. Interaction Between Program Coordination and Principal's Time	.05	1/21	2.05
Full Model (a, b and c)	.46	3/21	5.86**
Residual	.54		

*Significant at $\alpha \leq .05$.

**Significant at $\alpha \leq .01$.

Table 85. Intercorrelations Among Predictor and Outcome Variables in the 'Best' Model of School Average Gain in Student Achievement

	Variable			
Variable	1.	2.	3.	4.
1. School Average Gain	-			
2. Program Coordination	.53	-		
3. Principal's Time in Instructional Management	.34	0	-	
4. Interaction of (2) and (3)	.56	.38	.90	-

Several correlations in Table 85 are interesting. First, the zero correlation between the two main variables should be noted; principals were just as likely to devote more time to instructional management in schools with less coordinated programs as in schools with more coordinated programs. Secondly, the correlation of the interaction variable with school average gain is higher ($r = .56$) than the correlations of either of its component variables (program coordination and principal time) with gain.

Despite these suggestions of an incipient interaction effect, a sizable amount (40 percent) of the variance in achievement gains can be accounted for by a restricted model that ignores the interaction term, i.e., that hypothesizes a uniform effect of resource use over different levels of program coordination, and vice versa. The proportion of variance accounted for by a restricted model is statistically significant ($F = 7.77$, $\alpha \leq .01$).

4. SUMMARY AND DISCUSSION

A reanalysis of the 1975-76 ESAA in-depth resource-use data identified two distinctive profiles of school success that depended on the extent to which instructional programs were coordinated throughout the school. Among schools with highly coordinated programs, success in improving students' basic achievement was associated with greater commitment of the principal's time to program planning, curriculum development, and inservice training in program development and management. Among schools with less coordinated programs, success was associated with three classroom resources used in reading and math instruction: equipment, instructional aides, and the teachers' time. The identification of this latter pattern is very tentative, however, since only two in-depth study schools with less coordinated programs were successful.

Findings concerning one other resource, teacher inservice training, also indicated that program coordination was a key delineating feature of the two profiles of school success. Although somewhat more inservice training was associated with success in both types of schools, the impact was evident at significantly different resource levels, and the training was provided by different personnel. Teachers received less inservice training in schools with highly coordinated programs, and that which was provided was conducted primarily by on-site personnel--the principal and other teachers. By contrast, teachers spent significantly more time in inservice training in schools with less coordinated programs, and most of that training was provided by district employees, consultants, and other outside experts.

Finally, several models of school average gain in student achievement were analyzed, and these also pointed to the importance of program coordination. A model that used three predictor variables, (1) the extent of program coordination, (2) the overall cost of the principal's time in instructional management and, (3) a variable representing the interaction between these two, accounted for 46 percent of the variance in school average gain. Program coordination was the best single predictor, accounting alone for 29 percent of the variance.

Our reananlysis of resource data dealt with only a very limited set of the many factors generally observed to operate in schools. The remainder of this section presents a brief and somewhat speculative discussion of two other factors that may partly explain how and why program coordination conditions the impact of resource-use on school success. The two factors we discuss are educational philosophy and organizational work relations, that is, how authority and responsibility are allocated among teachers and administrators.

To some extent, differences in program coordination may represent differences in district and school policy, which in turn reflect different educational philosophies. One philosophy rationalizes an organizational structure in which teachers have a great deal of individual responsibility and authority for instruction. Teaching is thought of as an art or talent that finds expression through individuals. It may be aided by providing teachers with good tools and materials, but would be stifled by 'committee' procedures. Another philosophy holds that since learning proceeds sequentially, by steps that can be more or less well defined, then it can be facilitated by a proper organization of these steps. 'Good teachers' are sought in one case, and 'good teaching' in the other.

Corresponding to these different philosophies are two contrasting organizational styles, or ways of organizing work relations. Within the more traditional organization, teachers work largely within the confines of their own classrooms, quite autonomously, despite the superior authority of principals. A more recent approach has been to involve teachers more extensively in school policy decisions, and to structure teaching so that they are more involved with one another.

Associations among these three factors--program coordination, educational philosophy, and organizational work relations--may partly explain the conditioning impact of level-of-coordination on student achievement. One pattern of associations we might expect would be among (1) a high degree of program coordination, (2) a belief that learning is mostly the result of proper teaching methods and, (3) an organizational style in which there is much interaction and sharing

among teachers. A second pattern of associations would be (1) less program organization, (2) a philosophy that attributes learning largely to teachers rather than to techniques and, (3) an organizational style in which teachers work relatively autonomously.

Assuming that level-of-coordination has these relationships with educational philosophy and organizational patterns, it seems reasonable that students might benefit from different types of resources in highly coordinated programs than they benefit from in less coordinated programs. For example, we mentioned earlier that aides and equipment would seem amenable to the efficient use and management by individual teachers; aides may be easily accommodated in a structure in which teachers have most of the authority for what takes place in their own classrooms. But a different organization may be more effective in the use of other types of support staff--psychologists, compensatory teachers and other specialists--whose employment in schools is relatively recent and not yet widespread.

An indication that the way work is organized in schools may condition the impact of resource use on success is provided by the original 1975-76 ESAA in-depth resource use analyses, in which the use of remedial specialists was found to be more versatile among successful schools (Wellisch et al., 1977). Versatility was defined as the number of ways teachers could make use of a specialist: possible ways included calling the specialist in to help during reading or math, sending students to the specialist at any time, and regularly conferring with the specialist on methods and materials. While such use is not possible below some minimum of specialist time, organizational factors may be generally more critical to versatility--such factors as how authority and responsibility are delegated, shared and apportioned, and what kinds of enabling procedures are built into work relations. A compatible educational philosophy may also be important.

Research on these factors--on educational philosophy as expressed through school policies, on work relations, on lines of authority and communication among school staff--should provide more specific knowledge of the circumstances under which

schools may expect adequate returns on investments in various resources, as well as additional evidence that the way in which resources are used affects their impact on school success in raising student achievement.

Table 86. Intercorrelations Among Resources Representing the Principal's Time in Instructional Management Activities

Variable	Variable 1.	2.	3.
1. Program Planning, Needs Assessment and Program Evaluation	-		
2. Curriculum Development	.69	-	
3. Inservice Training in Areas of Instructional Management[a]	.61	.76	-

[a]Areas were: curriculum for disadvantaged students; needs assessment; program planning; and program evaluation.

Table 87. Average Per-Pupil Cost of Remedial Specialists Assisting Teachers With Primary Responsibility in Reading and Math Instruction

		School Success		
		Nonsuccessful	Successful	
Program Coordination	Low	$73	$69	$72
	High	$78	$83	$81
		$75	$80	$77

Table 88. Average Per-Pupil Cost of Speech Therapists Assisting Teachers with Primary Responsibility in Reading and Math Instruction

		School Success		
		Nonsuccessful	Successful	
Program Coordination	Low	$10.83	$17.41	$12.14
	High	$ 7.34	$20.42	$14.97
		$ 9.49	$19.75	$13.68

Table 89. Average Per-Pupil Cost of Media Specialists Assisting Teachers With Primary Responsibility in Reading and Math Instruction

		School Success		
		Nonsuccessful	Successful	
Program Coordination	Low	$5.80	$14.03[a]	$7.45
	High	$3.66	$ 5.20	$4.56
		$4.97	$ 7.16	$5.87

[a]While this mean appears high, the two schools on which it was based had costs at the extremes of the distribution. Using the Mann Whitney U Test, the difference is therefore not significant.

Table 90. Average Per-Pupil Cost of Counselors and Psychologists Assisting Teachers With Primary Responsibility in Reading and Math Instruction

		School Success		
		Nonsuccessful	Successful	
Program Coordination	Low	$1.25	$1.22	$1.24
	High	$2.06	$1.31	$1.62
		$1.56	$1.29	$1.45

NOTE: The above figures do not reflect the fact that in nearly 2/3 of the schools no help from counselors or psychologists was reported. The following table may therefore be of interest.

Table 91. Distribution of Schools Providing Counselors or Psychologists By School Success (With Mean Costs)

		School Success		
		Nonsuccessful	Successful	
Counselor or Psychologist	No	10 ($0)	4 ($0)	14 ($0)
	Yes	3 ($6.75)	5 ($2.33)	8 ($3.96)
		13	9	22

Table 92. Intercorrelations Among Predictor and Outcome Variables Used in Models of School Average Gain in Student Achievement

Model (identified by resource)	Variables	Intercorrelations			
Principal's Time in Program Planning, Needs Assessment and Program Evaluation (See text, Table 3)		1.	2.	3.	4.
	1. School Average Gain	-			
	2. Program Coordination	.53	-		
	3. Resource	.33	0	-	
	4. Interaction	.59	.51	.83	-
Principal's Time in Curriculum Development (See text, Table 4)		1.	2.	3.	4.
	1. School Average Gain	-			
	2. Program Coordination	.53	-		
	3. Resource	.29	0	-	
	4. Interaction	.49	.34	.92	-
Principal's Time in Inservice Training in Developing and Managing Instructional Programs (See text, Table 5)		1.	2.	3.	4.
	1. School Average Gain	-			
	2. Program Coordination	.53	-		
	3. Resource	.31	-.05	-	
	4. Interaction	.47	.23	.95	-
Teachers' Time in Program Planning (See text, Table 7)		1.	2.	3.	4.
	1. School Average Gain	-			
	2. Program Coordination	.53	-		
	3. Resource	-.38	-.11	-	
	4. Interaction	.04	.60	.72	-
Equipment Used During Reading and Math Instruction (See text, Table 8)		1.	2.	3.	4.
	1. School Average Gain	-			
	2. Program Coordination	.53	-		
	3. Resource	-.08	-.35	-	
	4. Interaction	.24	.18	.82	-
Instructional Aides Used During Reading and Math Instruction (See text, Table 9)		1.	2.	3.	4.
	1. School Average Gain	-			
	2. Program Coordination	.53	-		
	3. Resource	.02	-.20	-	
	4. Interaction	.14	.06	.94	-

Table 92. continued:

Model (identified by resource)	Variables	Intercorrelations			
Teachers' Time in Reading and Math Instruction (See text, Table 10)	1. School Average Gain	1.	2.	3.	4.
	2. Program Coordination	.53	-		
	3. Resource	.13	-.12	-	
	4. Interaction	.50	.79	.49	-
Teachers' Time in Inservice Training (See text, Table 11) NOTE: This set of intercorrelations is based on 22 schools.	1. School Average Gain	1.	2.	3.	4.
		-			
	2. Program Coordination	.48	-		
	3. Resource	-.02	-.06	-	
	4. Interaction	.22	.49	.83	-
Percentage of Inservice Delivery Time Attributed To On-Site Personnel (See text, Table 13)	1. School Average Gain	1.	2.	3.	4.
		-			
	2. Program Coordination	.53	-		
	3. Resource	.27	.63	-	
	4. Interaction	.35	.76	.97	-

Table 91. Analysis-of-Variance: Models of School Average Gain in Student Achievement, Each Using Three Predictor Variables: Program Coordination, A Resource Use Cost and a Variable Representing the Interaction of the Two

Model (identified by resource)	Source of Variation[a]	Proportion of Variance Explained	df	F value
Principal's Time in Program Planning, Needs Assessment and Program Evaluation (Table 3)	(1) Full model	.44	3/21	5.39*
	(2) Restricted model (additive)	.39	2/21	7.19*
	(a) Program Coordination	.28	1/21	10.37*
	(b) Resource	.11	1/21	4.04
	(3) Interaction	.05	1/21	1.80
	(4) Residual	.56		
Principal's Time in Curriculum Development (Table 4)	(1) Full model	.41	3/21	4.78**
	(2) Restricted model (additive)	.36	2/21	6.41*
	(a) Program Coordination	.28	1/21	9.81*
	(b) Resource	.08	1/21	3.00
	(3) Interaction	.04	1/21	1.52
	(4) Residual	.59		

[a] In every case, program coordination (2a) has been adjusted for the resource, and the resource (2b) has been adjusted for program coordination

*Significant at $\alpha \leq .01$.

**Significant at $\alpha \leq .05$.

Table 92. continued:

Model (identified by resource)	Source of Variation[a]	Proportion of variance Explained	df	F value
Principal's Time in Inservice Training in Developing and Managing Instructional Programs (Table 5)	(1) Full model	.42	3/21	5.04*
	(2) Restricted model (additive)	.39	2/21	7.01*
	(a) Program Coordination	.29	1/21	10.63*
	(b) Resource	.11	1/21	3.97
	(3) Interaction	.03	1/21	1.12
	(4) Residual	.58		
Teacher's Time in Program Planning (Table 7)	(1) Full model	.39	3/21	4.56**
	(2) Restricted model (additive)	.38	2/21	6.66*
	(a) Program Coordination	.24	1/21	8.27*
	(b) Resource	.11	1/21	3.66
	(3) Interaction	.01	1/21	.38
	(4) Residual	.61		
Equipment Used During Reading and Math Instruction (Table 8)	(1) Full Model	.31	3/21	3.09**
	(2) Restricted model (additive)	.29	2/21	4.41**
	(a) Program Coordination	.29	1/21	8.66*
	(b) Resource	.01	1/21	.42
	(3) Interaction	.01	1/21	.43
	(4) Residual	.69		
Instructional Aides Used During Reading and Math Instruction (Table 9)	(1) Full model	.30	3/21	2.94
	(2) Restricted Model (additive)	.29	2/21	4.38**
	(a) Program Coordination	.29	1/21	8.75*
	(b) Resource	.02	1/21	.47
	(3) Interaction	--	--	--
	(4) Residual	.70		
Teachers' Time in Reading and Math Instruction (Table 10)	(1) Full model	.35	3/21	3.71**
	(2) Restricted model (additive)	.31	2/21	5.06**
	(a) Program Coordination	.30	1/21	9.61*
	(b) Resource	.04	1/21	1.19
	(3) Interaction	.03	1/21	1.02
	(4) Residual	.65		
Teacher's Time in Inservice Training (Table 11)	(1) Full model	.25	3/19	2.11
	(2) Restricted model (additive)	.23	2/19	2.87
	(a) Program Coordination	.23	1/19	5.73**
	(b) Resource	--	--	--
	(3) Interaction	.02	1/19	.58
	(4) Residual	.75		

[a] In every case, program coordination (2a) has been adjusted for the resource, and the resource (2b) has been adjusted for program coordination.

* Significant at $\alpha \leq .01$.

** Significant at $\alpha \leq .05$.

Table 92. continued:

Model (identified by resource)	Source of Variation[a]	Proportion of Variance Explained	df	F value
Percentage of Inservice Delivery Time Attributed to On-Site Personnel (Table 13)	(1) Full model	.29	3/71	2.82
	(2) Restricted model (additive)	.29	2/21	4.20**
	(a) Program Coordination	.21	1/21	6.33**
	(b) Resource	--	--	--
	(3) Interaction	--	--	--
	(4) Residual	.71		

[a]In every case, program coordination (2a) has been adjusted for the resource, and the resource (2b) has been adjusted for program coordination.

*Significant $\alpha < .01$

**Significant $\alpha < .05$.

B. DEVELOPMENT OF CAUSAL MODELS RELATING SCHOOL AND CLASSROOM FACTORS TO STUDENT ACHIEVEMENT GAINS

This section deals with analyses of classroom characteristics and activities and their relationships with school characteristics, policies and practices, and with student achievement gains. Data on classroom characteristics came largely from teacher responses to questionnaires. These data consist of teacher demographic characteristics and attitudes, and classroom ethnic composition. Reports by ESAA in-depth observers constituted the chief source of data on classroom activities. Despite the care taken to train observers, it appeared that some differences in observational data might be ascribed to the observers themselves rather than to the events they observed. In general, observer ratings are notoriously subject to problems of limited inter-judge reliability. Consequently, preliminary analyses were performed to determine whether there might be systematic differences among observers' reports. There were two substantive considerations to these analyses: first, we wanted to determine what effects any observer-based rating differences might have had on the earlier in-depth analyses of classroom activities; second, we wanted to identify and minimize possible errors arising from such differences prior to conducting additional analyses in which observations of classroom activities would be central. Accordingly, we first present the results of the analysis of observer reliability.

1. OBSERVER RELIABILITY

The in-depth study conducted in 1975-76 utilized classroom observational data on reading and mathematics instruction collected in 176 classrooms in 26 in-depth elementary schools. Initial analyses of those data indicated that teaching practices during math lessons were more closely related to math achievement gains than the reading practices were to reading gains. These results paralleled those of other studies which had indicated that schools can have a greater impact on math than on reading achievement. In the light of our own and previous results, we elected to concentrate subsequent analyses upon data collected for math instruction only. Thus, the present discussion deals with observational data collected by 12 observers during math lessons in in-depth classrooms.* Each

*No math instruction was observed in 51 of the total sample of the classrooms visited.

observer visited and recorded data in from 7 to 14 classrooms; only one observer visited any given classroom. In each classroom, the observer recorded data on a total of 45 different observational items. Some items called for counts, and others for estimates of the frequency of particular teacher and student behaviors. Still other items called for more global assessments of teaching strategies, teacher demeanor, or classroom organization.

a. Identifying Sources of Observer Bias*

Though there are potentially a large number of systematic types of observer bias, we elected to investigate one of the simplest types--the type that occurs when an observer over- or underestimates all observations of a particular item by a constant amount. We investigated this type of bias because many of the observational items were estimates of the extent or frequency of a type of behavior. Also, we were able to demonstrate that a simple statistical model for estimating the magnitude of this type of bias could be applied to our data. A discussion of that model and its applicability to these data follows.

When classrooms are randomly assigned to observers, an appropriate statistical model for estimating the magnitude of observer bias which is uniform across all observations is the analysis-of-variance. Under the random-assignment assumption, the expected value of the observer mean on each item is equal for all observers. The presence of observer bias is indicated by differences in item means between observers. The test for the significance of these differences is the ratio of the between-observers to within-observers mean squares. However, in the in-depth study, the situation regarding the estimation of this type of

*In the following discussion we refer to tendencies of individual observers to give different ratings for observations of the same data as 'bias'. We have only indirect evidence on such differences, since only one observer visited any given classroom. The methodological problem that results when such observations are combined to form variables, we refer to as 'unreliability'. We recognize that in the case of individual ratings our nomenclature is unsatisfactory. Using the term 'bias' for the occasion when one observer would record 'frequently', another 'sometimes', does violence to our everyday language. Still, the term is preferable to its statistical equivalent, 'error', which has an even more absolute connotation in everyday use. The reader will, of course, recognize that some such term is necessary in our discussion of the statistical model employed here.

bias was not clear because the assignment of classrooms to individual observers was not strictly random. For this reason we approached the estimation of observer bias effects in the more indirect fashion described below.

Although we knew the assumption of random assignment was not met by our data, we first examined these data using the analysis-of-variance model. Results from the breakdown of the 45 math observational items by sets of classrooms linked with observers showed significant (.05 level) observer effects for 62 percent of the variables (28 of 45). Since this is considerably greater than the five percent expected on the assumption of random assignment, we concluded that our observational items showed prima facie evidence of observer bias. However, we felt it was then necessary to try to determine whether these results were, in fact, due more to nonrandom assignment of observers to classrooms than to observer bias.

There were several reasons for concern about violations of the random assignment assumption. The 1975-76 ESAA in-depth study observers visited the 26 elementary schools in two-person teams, dividing the class observation work at each school as they saw fit. Such a procedure clearly departs from random assignment, and is likely to introduce several sources of bias in addition to those we were interested in. Another procedure that led to non-random assignment of observers to classrooms was the sending of teams to several schools in close proximity, for logistical reasons. In consequence, regional variation in teaching practices would tend to be reflected in our analysis of between-observer effects. Finally, a nonrandom assignment of observers to classrooms was inevitable as a result of sending the teams to a selection of schools rather than classrooms. Most observers visited several classrooms in 4 or 5 of the 26 schools, and zero classrooms in the remaining 20 or so schools. As a result, any tendency for school factors (such as school-philosophy of instruction) to be reflected in classroom teaching practices would also tend to contaminate our analysis.

In spite of these possibilities, three convergent lines of evidence led us to conclude that the sets of classrooms observed by particular observers were not systematically different from each other in any important way, and consequently

the departures from random assignment were not so serious as to make the analysis-of-variance model inappropriate.

The first line of evidence on this point came from an examination of the 119 questions concerning attitudes and philosophy of instruction which teachers from each classroom answered. In order to determine whether the sets of teachers (each set associated with the classrooms visited by a particular observer) tended to differ from each other in attitudes and philosophy of instruction, an analysis-of-variance of the responses to these 119 questions was performed. Eight of the 119 questions (6.7 percent) showed differences in between-observer means that were significant at the .05 level. Since we would expect five percent of these differences to be significant if teachers were assigned to observers on a strictly random basis, we concluded that in this respect, at least, the data exhibited a relatively close fit to the random assignment model.

A second type of evidence consistent with the random model came from a parallel analysis-of-variance of student achievement gains. Teachers, grouped by observers, did not differ in the achievement outcome gains experienced by their students. The F ratios for between to within-set variance were .42 for math, and .92 for reading gain. We regard those results as particularly important, because 'real' differences in teaching practices--those not due to the result of observer bias--ought to be reflected in student outcomes.

A third line of evidence ruled out the possibility that between-observer effects were significantly confounded with school effects. We were concerned about that particular possibility because each observer gathered data in several classrooms in each school. To estimate the magnitude of these effects, we reexamined the 28 observational items for which there were significant between-observer differences. The data were repartitioned into schools, and observers within schools. Next, we ran an analysis-of-variance on this nested design. The between-school variance was not significant in comparison to the observer variance within schools for any of those items; consequently, the effects were attributed to between-observer effects rather than school differences for the 28 items.

Together the three lines of evidence recounted above convinced us that departures from the random assignment model were not serious. In consequence, we felt that the 28 significant between-observer effects on observational items were due to observer-introduced bias. Eleven of these 28 items were eliminated from further analysis because there was evidence of unsystematic bias as well; that is, comments by observers during debriefing strongly suggested a lack of common understanding of these items. The discussion that follows concerns our efforts to eliminate or adjust for bias in the remaining 17 items that had significant evidence of observer bias.

(1) Identifying Outlier Observers

Our first step in dealing with the problem of observer bias was directed toward determining whether the effects we had located were due to a small number of deviant observers. We reasoned that if we could identify a few observers who were scoring items markedly different from the others, then we might eliminate observer bias simply by excluding those observers from subsequent analysis.

We attempted to identify deviant observers by using a discriminant function analysis in which the groups were sets of classrooms associated with particular observers, and the predictors were the 17 observational items with significant observer effects. The extent to which an observer's mean values on the discriminant functions deviated from the grand mean indicated the observer's overall outlier status. Using this technique we were able to identify two outlier observers. The utility of this method was confirmed when we examined the instruments used by these two observers. We found that their response patterns and coding procedures were so idiosyncratic as to justify the elimination of their observations from all further analyses. Eliminating the observations associated with the two outlier observers reduced the number of classrooms in our sample from 125 to 109. We then re-ran the analysis-of-variance of between-observer effects, finding that 11 of the 17 observational items still showed significant observer effects. Based on those results, we concluded that excluding a small number of cases associated with outlier observers was not sufficient to eliminate observer bias. In consequence, we examined additional means of correcting the bias.

(2) Identifying Bias Associated With the Race and Sex of Observers

Another approach to the bias problem arose out of the recognition that different types of observers might tend to rate observations in systematically different ways. We thought it reasonable, for instance, that the sex and ethnicity of observers might be factors in observer bias. To investigate this possibility, we ran a two-way analysis-of-variance for effects of the race* and the sex of observers. Six of the 17 observational items showed significant (.05 level) effects for sex, one item showed significant effects for race, and there were four significant two-way interaction effects.

Next, we examined the direction of the race and six differences for all 17 items. For the subset of 9 items, where some favorable or unfavorable evaluation of teaching practice seemed to be involved, we were able to locate a pattern. Black observers (7 of 9 items) and male observers (6 of 9 items) tended to make evaluations that reflected favorably on teaching practices. These results (displayed in Table 93) are consistent with those connections between observer characteristics and bias that might reasonably be expected. For instance, it seems likely that the differences associated with the ethnicity of observers may be related to the different school experiences of the observers when they themselves attended school. The kind of process that may be operating has been discussed in another context by Wallen and Travers (1963). Commenting on the strength of school background experiences as compared with formal training as a factor in teaching practices, those authors observed that most teachers "...teach in accordance with the pattern they observed when they were pupils", an experience that "...provides a much more vivid guide...than does the [relatively short] period of teacher-training...". Similarly, observers might be expected to draw upon their years of experience as students. Just as with teachers, it may be expected that the many years observers spent as students provided them with 'a much more vivid guide' to perceiving the nature of classroom activity than did the two-week training period. In this regard it seems likely that blacks, as children, probably had less satisfactory school experiences than their white counterparts. Either they went to low quality segregated schools, or they were the least

---

*Black vs. other

Table 93. Observer Sex and Ethnicity, and Evaluations of Teaching Practices*

(Plus (+) indicates a more favorable evaluation; minus (-) indicates a less favorable evaluation.)

Items	Blacks	Males
Tasks are clearly presented to the students.	+	+
Teacher usually gives feedback on the adequacy of student responses.	+	+
Teacher monitors how well students are learning	-	+
Teache frequently offered constructive criticism in response to inadequate answers.	+	-
Teacher seldom used harsh criticism or ridicule in response to inadequate answers.	+	+
Teacher seldom lectures (scolds) class.	+	-
Teacher seldom yells or shouts.	-	-
Teacher is task oriented.	+	+
Teacher is warm, supportive.	+	+

*Several of these items appear again later in this section in the discussion of classroom factors related to Student Task Articulation. The third item in Table 93 is referred to as Monitoring of Students; the next to last item as Task Orientation. The three items dealing with harsh criticism, lecturing, yelling or shouting are combined in the index Avoidance of Harsh Discipline.

favored in desegregated schools. As adult observers, they may be expected to have lower expectations regarding educational excellence, since their childhood experience did not provide it.

The reason for sex-linked differences in female observer judgments is less clear, but our personal acquaintance with the observers suggests that they may have identified more closely with teachers than their male counterparts. We think the closeness of the identification may have encouraged a finer attention and more critical attitude toward classroom teaching practices than was the case for male observers.

b. Adjustments Made for Bias Associated With the Race and Sex of Observers

The direction of these biases reinforced our belief that differences in observers' race and sex might account for between-observer effects. In consequence, we decided to compute adjusted scores for each of the 17 items with significant observer effects. For each observation, the adjustment was equal to the difference between the mean calculated for groups of observers classified by race and sex and the grand mean calculated for all observers.

Next, we re-calculated the between-observer effects for each of the 17 items, using the adjusted scores. In this reanalysis, between-observer effects were significant for only four of the 17 items that had shown significant effects when unadjusted scores were used. We regarded these adjustments as relatively more successful in reducing observer bias than the exclusion of outlier observers alone.

In another respect, however, these adjustments proved less useful; they only marginally improved the prediction of class-level math gain. The correlations of adjusted and unadjusted items with math gain are displayed in Table 94. As is evident in the table, the improvements in the correlations of the adjusted items with math gain are small. The adjustment has the effect of eliminating a significant correlation for one item: amount of student nontask interaction. However, there are three items ('Tasks are clearly presented to students', 'Teacher seldom lectures (scolds) class', Teacher is task-oriented') which have

Table 94. Comparisons of Original and Adjusted Items

Items	Correlation With Math Gain		f ratio	
	Original Item (with outliers)	Adjusted Item	Original Item (with outliers)	Adjusted Item
Teacher provides context for lesson with lesson introduction.	.30*	.30*	3.16*	1.96
Tasks are clearly presented to students.	.13	.19*	2.64*	1.42
Teacher usually gives feedback on the adequacy of student responses.	-.03	-.07	3.26*	3.41*
Teacher monitors how well students are learning.	.03	-.01	4.78*	1.98
Teacher uses constructive criticism.	.01	-.01	2.64*	1.00
Teacher uses harsh criticism/ridicule.	-.13	-.13	1.97*	.84
Amount of student non-task interaction	-.18*	-.14	2.91*	1.44
Average proportion of students task oriented	.24*	.24*	2.22*	1.90
Teacher seldom lectures (scolds) class.	.13	.18*	2.52*	1.07
Teacher is task-oriented.	.14	.20*	3.58*	1.07
Teacher is warm, supportive.	.13	.14	2.15*	.54
Percentage of lesson period students use programmed math text	.00	-.02	3.03*	1.78
Percentage of lesson period students use non-text books	-.09	-.13	7.50*	7.48*
Percentage of lesson period students use math puzzles, games, toys	-.11	-.12	3.07*	2.70*
Percentage of lesson period students are tutored on a one-to-one basis	-.09	-.13	3.43*	2.22*
Percentage of lesson period students spend working independently	-.02	.04	3.57*	1.39
Percentage of lesson period spent in practicing	-.08	-.14	7.10*	1.54

*Significant at the .05 level

significant correlations with gains ($P \leq .05$ in the expected direction), but only in their adjusted versions. In view of these improvements and reductions in the size of observer effects, we elected to use adjusted items in subsequent analyses of classroom activities. Those analyses, using adjusted items and variables constructed from adjusted items, are reported in the immediately following section.

2. THE RELATIONSHIP OF SCHOOL AND CLASSROOM FACTORS TO STUDENT ACHIEVEMENT GAINS*

In this section we describe analyses which attempted to connect classroom characteristics and activities both to school factors and student math achievement gains. Our discussion begins with analyses directed at the question of whether different sets of classroom factors were associated with student gains among subgroups of students. We were interested in this question because of the possibility that earlier analyses might have failed to identify some classroom factors, either characteristics or activities, which are effective with some types of students but not others. If such factors could be found, then we would have a basis for constructing more than one model of effective schooling for subgroups of students. We discuss the data bearing on this question in the immediately following subsection. After resolving the question of whether one or several models are required, we discuss our analyses of the relationships of school factors to classroom factors.

a. Classroom Factors and Student Math Achievement Gains in Student Subgroups

The first set of analyses examined the relationships of a large number of classroom factors to student math achievement gain in different student subgroups. One line of investigation was directed at determining whether teachers reinforced white and minority students differently, as, for example, by applying lower standards for praising minority students. Other investigators had hypothesized that teachers with low standards praised minority students gratuitously for substandard work (Fernandez et al., 1975). This they felt accounted for the slower progress of minority students. Basing our analyses on student reports of teacher reactions to their work, we did not find evidence to support this hypotheses. In our sample, the math gains of minority students were only marginally lower than white students at equivalent pretest levels. Also, student reports of teacher reinforcement practices were unrelated to academic student progress both in white and minority subgroups.

---

*The analyses reported here were conducted at three levels (school, class, and student) and included 25 in-depth schools, 109 classrooms, and 2,004 students. (It should be noted that one of the original schools was dropped because of poor data quality.) In these analyses, only math gains are dealt with since earlier studies revealed a much stronger relationship between classroom factors and math gains, as opposed to reading gains.

Next, we examined the within-subgroup correlations (correlations calculated separately for whites and minority students, and for 3rd, 4th, and 5th-grade students) between a large set of classroom factors and residualized math gain.* We were most interested in the thirty-plus variables obtained from classroom observations, describing classroom activities. We also examined over twenty additional variables, obtained from questionnaires and interviews, describing teacher and classroom characteristics. By and large, subgroups did not differ appreciably in their correlations between classroom factors and math gain. However, we did test the significance of the largest of these differences, using a regression approach.** There were no significant subgroup interactions, and no new variables related to math gain were identified. Four of the five variables that had been the best predictors of math gain in earlier classroom level analyses were significant predictors of residualized gain in this student-level analysis.*** Though there were no significant interactions, one set of relationships seems worth noting here; this involves the associations between subgroup gains and teacher beliefs about the importance of two student characteristics, race and family income, in explaining student academic success and failure. Achievement gains of minority, but not of white students, appeared to be related to the beliefs of their teachers about the importance of race and family income in determining student academic success. Minority students with teachers who believed race and family income were unimportant had higher math gains. Math gains for

---

*The residual deviation of posttest score from predicted posttest score (from the regression equation of math posttest on pretest).

**First, we regressed residualized gain upon a variable representing the classroom activity (teacher characteristic or behavior) and upon a dummy variable representing student ethnicity or grade level. Next, we added a variable representing the interaction of student and teacher on classroom characteristics, i.e., the cross-product of the two variables. If the additional contribution of any of these interaction variables to the total variance explained by the regression equation had been significant, we would have concluded that some inputs were more effective in particular subgroups.

***These four variables were based on observations and teacher reports that: (1) the day's lesson was introduced in relation to the previous lesson; (2) the teacher made extensive use of positive reinforcement and had beliefs favorable to the use of reinforcers for academic progress; (3) the teacher favored setting challenging goals and the use of behavioral objectives; (4) students were largely engaged in task related behavior. A full discussion of the variables constructed from these reports and observations is found in Wellisch et al., 1977.

white students were unrelated to teacher ratings of the importance of these factors: $r = .12$, $\alpha \leq .001$ and $r = .08$, $\alpha \leq .01$ for minority students; $r = .01$, NS and $r = .03$, NS for white students. The differences between the two correlations are small and not statistically significant. However, they are in a direction consistent with a reasonable interpretation of the effects of teacher attitudes on student subgroup gains. We may assume that teachers who think student race is an important factor in student academic success believe that minority students suffer from a racial handicap. Low family income is also a characteristic commonly thought to be a minority handicap. On the other hand, teachers who think these student attributes less important presumably are more optimistic about the influence of other factors, including their own inputs, to minority student gains. Under these assumptions we might speculate that teachers of this optimistic opinion may make greater efforts to assist educationally disadvantaged minority students, while teachers of a contrary opinion would assist these students less, and ascribe the students' lack of progress to factors outside the teacher's control.

3. THE RELATIONSHIP OF SCHOOL TO CLASSROOM FACTORS

The results of the subgroup analyses reported above indicated that it would not be necessary or productive to construct separate models of effective schooling for student subgroup populations. Consequently, we turned to an examination of the relations between classroom activities and school characteristics, policies, and practices. The major concern in these analyses was to interpret some rather strong school-level relationships that had emerged from our initial examination of the second-year data (1975-76). That examination had revealed five school organizational variables that were positively related to school success in raising student achievement: administrative concern for instruction, communication of an administrative point of view on instruction, administrative responsibility for instruction, school-wide coordination of instructional programs, and administrative policies concerning academic standards. Our primary objective in the new analyses with respect to these five variables was to identify their connections with classroom characteristics and activities, particularly those activities that were related to math gains.* One such connection was identified--that between school-wide coordination of instructional programs and the orderliness of students as they progressed from one task to another in their classwork. In the paragraphs that follow, we deal with that relationship first, after which we discuss some other findings which, though not specifically tied to math gain, are of interest on other grounds.

a. School Program Coordination and Student Progress From Task to Task

The single relationship between a school organizational factor (related to school success) and a classroom variable (related to math gain) was between school-wide coordination of instructional programs (called here School Program Coordination) and the orderliness of student progress from task to task (called here Student Task Articulation). School Program Coordination was measured by an index combining the principal's and teachers' assessments of the extent to

---

*We also examined the relations between a wide range of other school characteristics and classroom activities, since we thought these analyses might increase our understandings of math programs operations apart from questions of schools' success. We report some of the more interesting results of these analyses under the heading 'Other Findings'.

which the school's instructional program was coordinated school-wide. This index ranged from 0 to 4, with 4 indicating full consensus that the school program was coordinated "...to a large extent, both in many aspects and throughout all grades". A score of 0 indicated a consensus that the program was coordinated "very little". Earlier analyses showed that this index was positively and significantly associated with school success in raising math achievement ($\alpha \leq .03$, Mann Whitney U test). The second variable, Student Task Articulation, was measured by a three-point scale based on observations of two math lessons in each class. A score of 3 on this scale indicates that all or most students move from task to task "...comfortably...in an orderly and controlled fashion with few or no disruptions". A score of 1 on this scale indicates that there are many disruptions and that the teacher is "...largely occupied with controlling the class and maintaining order".*

In the student-level analyses reported here, Student Task Articulation was significantly related to residualized student gain, ($r = .12$, $\alpha \leq .001$). The association between School Program Coordination and Student Task Articulation was positive and significant ($r = .37$, $\alpha \leq .05$), indicating that students progressed from task to task in a more orderly manner where instructional programs were more highly coordinated.

Before presenting our interpretation of that relationship, we want to draw attention to a relation between another school organizational variable and Student Task Articulation. The school organizational variable we refer to came from principals' reports of the extent to which the principal and the teachers keep each other well-informed. That variable also showed a significant positive correlation with Student Task Articulation ($r = .34$, $\alpha \leq .05$), indicating students were more orderly in schools with better communication between the principal and the teachers. Though not significantly correlated, these two school variables do have a conceptual affinity. Communication is one means by which program coordination can be achieved.

---

*This item was a component of the Student Task Related Behavior Scale used in the already reported classroom-level analysis of factors related to math gains. In those analyses the scale was a significant predictor of mean classroom math gains.

Our interpretation of these relations can be represented as shown in Figure 25 a path diagram* in which the connection between School Program Coordination, Principal-Teacher Communication, and Student Task Articulation is mediated by the following five variables representing math teaching practices:**

- Task Orientation (teacher)
- Avoidance of Harsh Discipline
- Monitoring of Students
- Use of Math Instructional Objectives
- Individualized Mode of Instruction

In the paths from the five mediating variables to Student Task Articulation, the numerical values associated with the arrows are standardized regression coefficients. For the paths from School Program Coordination and Principal Teacher Communication to the five mediating variables the numerical values are zero-order correlation coefficients (equivalent to the standardized coefficients in the single-variable case).

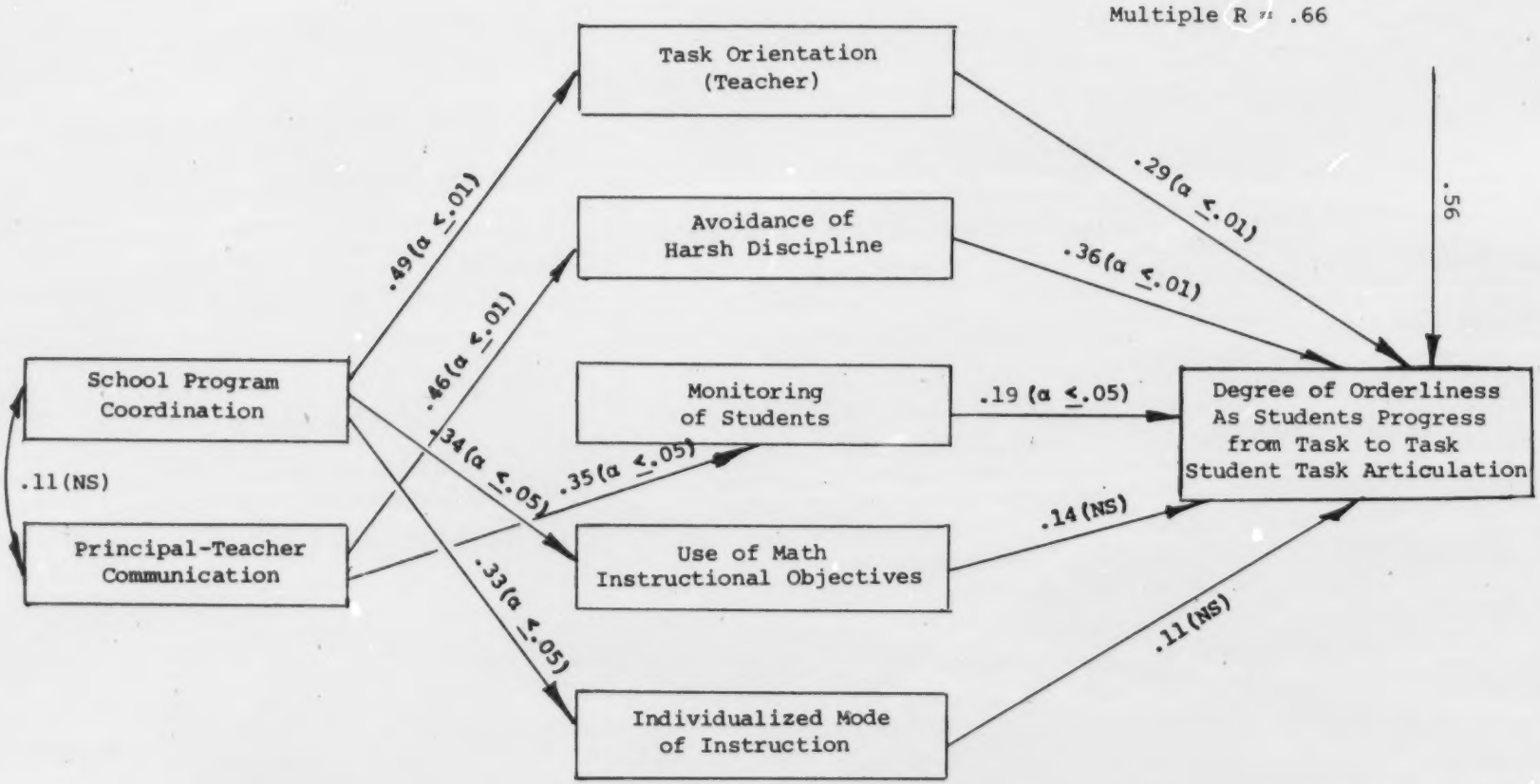
The path diagram shows no direct paths from either School Program Coordination or Principal-Teacher Communication to the Student Task Articulation variable. The reason for this is that when the mediating variables are included in a regression equation with the two school program variables, the direct effects of these latter variables are not significantly different from zero: .07 and .01,

---

*Our presentation departs from convention in that regression coefficients from two levels of analysis are shown. The paths from School Program Coordination and Principal-Teacher Communication to the five mediating variables are from a school-level analysis. The paths from mediating variables to the Student Task Articulation are from a classroom level analysis. The diagram represents the causal flow from level to level. However, using this representation, it is not possible to construct the matrix of school or classroom-level correlations by multiplying the regression weights in the diagram.

**Teacher Task Orientation, Monitoring of Students, and Individualized Mode of Instruction are single item indicators based on 3 and 4 point observer ratings. Avoidance of Harsh Discipline is an index constructed from the standardized observer ratings of the three items dealing with use of harsh criticism or ridicule, lecturing (scolding), and yelling or shouting. Use of Math Instructional Objectives is a dichotomous indicator from teacher reports that they have "...a record showing specific instructional objectives (in math) and the number of students who have attained each of them".

Figure 25. School Organizational Variables and Classroom Factors Related to Student Task Articulation*



*All nonsignificant paths except those from Use of Math Instructional Records and Individualized mode of Instruction have been eliminated. These two nonsignificant paths were retained because they facilitate interpretation of the connections between the two organizational variables and the degree of Orderliness as Student Progress from Task to Task.

respectively. That being the case, all of the effects of the two organizational variables are shown in the diagram as mediated by these teacher behaviors.

Three of the mediating variables in the diagram deal with classroom management activities, Task Orientation, Avoidance of Harsh Discipline, and Monitoring of Students. Classrooms where students made more orderly progress from task to task had teachers who made greater use of these three management techniques. Such classrooms were more commonly found in schools with coordinated instructional programs and good communication between the teachers and the principal. Specifically, teachers were more task oriented in schools with coordinated instructional programs. The remaining two effective classroom management techniques--Monitoring of Students and Avoidance of Harsh Discipline--were more common in schools with good communication between the teachers and the principal.

We interpret these associations as evidence that certain school management practices encourage teachers to run their classrooms efficiently. In the context of these associations we think it is reasonable to regard these effective classroom management activities as the logical extension of a well-communicated set of instructional objectives consistent with a coordinated school program.

The remaining two mediating variables are concerned with teaching methods. These are the Use of Math Instructional Objectives, and the Individualized Mode of Instruction. Classrooms where students made more orderly progress from task to task had teachers who made greater use of these two methods. Each of these methods was more common in schools with coordinated instructional programs. We believe these associations are supportive of the line of reasoning developed in the preceding paragraph. They suggest another means by which coordinated school programs might have an impact upon classroom activities to encourage student task-related behavior. Coordinated school programs were defined as those in which instructional activities were organized "...in many aspects (context, sequence of objectives, materials) and throughout all grades".* The use of

---

*From the Principal Interview.

Instructional Objectives in math and the Individualized Modes of Instruction are mechanisms for ensuring that each student has a task that is consistent with the school's instructional program and is also appropriate to the student's current achievement and abilities. Well defined objectives, and individualized instruction directed toward those objectives, might be expected to occur more frequently in an overall school program that is organized around a set of learning objectives, i.e., in a coordinated school program. It also seems reasonable that teachers and students might function more effectively in schools with coordinated programs. In such schools both teachers and students can behave more effectively because it becomes possible to know what is expected. One way these expectations are apparently conveyed is through effective principal-teacher communications. Teachers, in turn, convey these expectations through the use of individualized learning objectives, and by utilizing effective management techniques.

4. OTHER FINDINGS

The analyses reported above focus upon variables that were tied to math achievement gains. Those relationships were of special interest because they served to interpret associations between school organizational factors and school success. However, in addition to those results, we identified other relations of some interest, though not related to math gains. The most interesting of these relations concerned the school's racial/ethnic composition, and two classroom observation variables: the use of One-to-One Tutoring, and the Proportion of the Math Lesson Period devoted to practice. In schools with higher proportions of minority students, one-to-one tutoring occurred less often ($r = -.46, \alpha \leq .01$), and a greater proportion of the lesson period was devoted to math practice ($r = .36, \alpha \leq .05$). In schools with higher proportions there was a nonsignificant tendency to devote a greater proportion of the main lesson period to whole instruction ($r = -.28$, NS). These correlations may indicate that schools with high proportions of minority students emphasize less innovative instructional techniques, that is, they focus on whole-class instruction and practice, rather than on one-to-one tutoring. This is consistent with the fact that ESAA in-depth schools with high concentrations of minority students tended to be located in rural areas, isolated from the centers of educational innovation. And while the lack of innovativeness in these predominantly minority schools had no apparent adverse effect on school success, the predominantly minority schools were somewhat more successful in the In-Depth Study. These findings may be of some interest to developers of educational approaches targeted for the most disadvantaged populations. Many of those students may be located in schools with the least receptivity to innovation.

C. EFFECT ON IN-DEPTH STUDY RESULTS OF USING DIFFERENT DEFINITIONS OF SCHOOL SUCCESS

Good information about what actions and inputs result in greater student achievement is vital to responsible policy makers in education. Implicit in any decisions they approve are ideas of causality; for example, it might be felt that 'involved parents foster student growth', 'students learn more in smaller classes' or 'high teacher morale enhances student achievement'. However, research that is intended to provide the rationale for decisions is typically confined to correlational analysis, which is incapable, by itself, of supporting conclusions about causality. For example, the ESAA in-depth studies revealed a correlation between teacher morale and student achievement-level, but this finding is ambiguous about the direction of causality. Higher teacher morale might have caused the greater student achievement; alternatively, however, teachers may be more satisfied as a consequence of student achievement, or students may achieve more and teachers be happier, each as a result of a third factor--clean, tidy, and well-functioning facilities, perhaps.

The purpose of the research reported here was twofold. One was to assess the preponderant direction of causality between teacher morale and student achievement by comparing results that have been obtained using different outcome measures. Some previous studies have reported associations based on achievement level or gain that was measured before morale was assessed, while others have been based on achievement that was measured afterwards.* The preponderant causal direction--whether teacher morale affects student achievement more than achievement affects morale--was evaluated by comparing the strength of the associations of each type. The possibility that the relationship between them is spurious, or that a third factor largely accounts for the relationship between them, was also evaluated.

The second purpose concerned basic methodology and was based upon the results of the evaluation suggested by the first objective. Should comparisons show that the choice of the outcome substantially affects the results obtained, then the reputed 'impact' of 'inputs' in general should be carefully re-evaluated.

*The studies we refer to have all dealt with school-level variables. Therefore, throughout this report 'teacher morale' (or satisfaction) and 'student achievement' both refer to a collective or average measure in a school.

Consequently, the goal was to demonstrate and stress the fact that interpretations of results are constrained by the choice of the outcome measure.

This section is organized in two main parts, each presenting the results of a given approach to the problem. The first approach was to review the relevant literature and evaluate the findings on the relationship between teacher morale and student achievement, comparing them according to how the achievement measure was constructed. While there is substantial merit in reviewing a number of research efforts, this method is not ideal. Most research has not been explicitly designed to examine causality between these or any variables. Furthermore, studies differ in various ways, and there is no way of controlling for whatever may have been critical dissimilarities.

The second method of approaching the question depended on parallel analyses within a single study, and therefore did not suffer from uncontrolled dissimilarities. Two studies were available for which these parallel analyses were possible: the 1974-75 and the 1975-76 ESAA In-Depth Studies. In effect, each in-depth study was treated as three parallel studies, with each version precisely identical in every respect except for the outcome measure. Both included data that allowed construction of the relevant alternate version of achievement gain, as well as data on prior achievement level. Associations of teacher morale with the three outcomes--prior achievement gain, study-year achievement gain, and prior achievement level--were then compared.

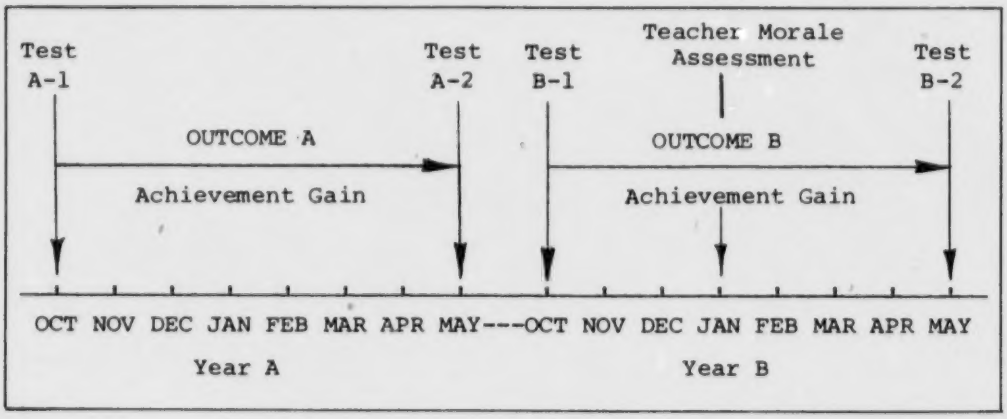
Certain considerations are important to bear in mind when examining causality among these variables. Before presenting the results, our assumptions about how the difference in outcome measures aids in evaluating the two major competing hypotheses, and the limitations that apply to this evaluation, will be discussed.

1. OUTCOME MEASURES AND THE TWO MAJOR COMPETING HYPOTHESES

The analyses to be reported here deal with only one dimension of difference between alternative measures of school success in raising student achievement: the period over which achievement gain takes place relative to the time at

which teacher morale is assessed. Figure 26 diagrams the ordering of the 'events', showing that in one case gain precedes the assessment of teacher morale (Outcome A), while in the other gain takes place largely after data on morale have been collected (Outcome B).*

Figure 26. Times at Which Achievement Gain and Teacher Morale Are Assessed



Each outcome measure corresponds to one of the two major competing hypotheses, diagrammed below in Figure 27. Arrows indicate the causal direction. Studies that use an outcome measure based on past gain (and others that use one based on prior achievement) implicitly and logically test the hypothesis, 'student achievement affects teacher morale'. This we have termed here 'Hypothesis A'. But because achievement is, a priori, the outcome, inferences in the opposite direction, 'morale affects achievement', are regarded as legitimate. This we have termed here 'Hypothesis B'. In fact, these studies are not generally considered longitudinal at all, but since achievement, and the factors that affect it, are assumed to be stable over time, an input/output model is applied even though output data are collected prior to input data.** It is assumed that

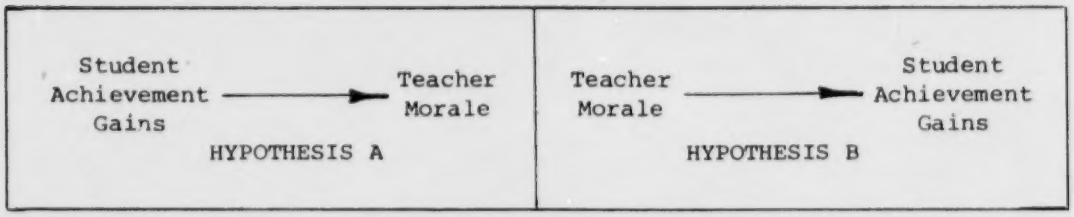
---

*In this discussion, the third 'outcome', prior achievement level (Test A-1), A-2 or B-1) is not differentiated from prior gain, since they are alike in the critical respect, the fact that both are prior measures. However, achievement level differs in other respects that will be discussed later.

**Some studies, particularly major evaluations, conceptualize 'input' as 'interventions' and 'output' as 'outcomes'. We have used 'output' and 'outcome' interchangeably.

schools that performed well academically in the past will continue to do so, and that schools that currently rate high in teacher morale rated high in previous years. Either hypothesis is obviously tenable under these assumptions.

Figure 27. The Two Major Competing Hypotheses Explaining A Relationship Between Student Achievement Gains and Teacher Morale



Other studies use an outcome based on study-year gains. Input variable data are collected at a time logically proper for use with an input/output model. The hypothesis 'teacher morale affects student achievement' (Hypothesis B), is implicit in the assessment of morale (the input) prior to achievement (the outcome). The two ESAA In-Depth studies were of this type.

2. ASSUMPTIONS AND LIMITATIONS OF THE ANALYSES

A basic assumption underlying our assessment of the preponderant direction of causality is that there is a longitudinal component in all these studies that should not be disregarded. Yet year-to-year stability of school success in raising student achievement is rarely examined. The ESAA In-Depth studies are among the few that have done so, and the results indicated that presumptions of stability in program characteristics and outcome levels may not be warranted.*

Year-to-year stability is generally discussed in relation to relatively short-term effects. It seems reasonable to suppose there are also long-term effects involving teacher morale and student achievement. However, because the longi-

*In the first ESAA In-Depth Study, half the reading classifications (12 of 24) and nearly two-thirds of the math classifications (15 of 25) changed from successful to nonsuccessful (or vice versa) from 1973-74 to 1974-75. More stringent criteria resulted in a somewhat more stable sample for the second study; even so, 8 of 25 schools changed classifications from 1974-75 to 1975-76.

tudinal component is weak, most studies are not well suited for dealing with complex, dynamic social phenomena over time, such as when the effects of a former condition persist even though the condition has ceased to operate, or when the effects of a new effort are evident only after it has been applied over some period.* For the same reason, these studies are not designed to test complex causal ordering such as feedback loops, e.g., as morale improves, achievement gains increase, which further improves morale, and so on. Constrained by the available data, our models do not include these factors.

Finally, since the purpose of this research was not to explain either teacher morale or school success in raising student achievement, but to evaluate the preponderant causal direction between them, the number of other variables considered was limited. Although we may be certain that many other variables affect both teacher morale and school success, only two seem important in this context. One variable, specifying when teachers are informed of test results, has theoretical importance to Hypothesis A. However, neither the ESAA studies nor others we reviewed include this variable; consequently, we assumed it to be an invariant intervening variable to Hypothesis A, and as such, it was not further analyzed.

The second variable we considered important was the socio-economic status of the student body (SES), particularly in evaluating relationships between prior achievement <u>level</u> and teacher morale. Although achievement <u>gain</u> may also be partly due to student background, relatively more of the variance in achievement <u>level</u> may be explained by non-school factors.** (Achievement level may be conceptualized as: current level = gain + initial level, where initial level is due entirely to background factors such as IQ, home training, and so on, and gain is due both to background and to school factors.) Our chief concern with

---

*The nature of certain variables may be misrepresented by the fact that data are collected at a point <u>in time</u> rather than <u>over time</u>. For example, in Figure 30 teacher morale appears as an event or intervention, whereas common experience suggests it is a condition that fluctuates.

**A relationship between student SES and student achievement has been established by many studies, most notably the 1965 U.S. Office of Education Equal Educational Opportunity Survey (the Coleman Report.)

student SES in this study was to test whether it does or does not relate to teacher morale and student achievement in such a way that the relationship between morale and achievement is completely spurious.

3. REVIEW OF LITERATURE ON RELATIONSHIPS BETWEEN TEACHER MORALE AND STUDENT ACHIEVEMENT

Following World War II, a large body of research on schools focused on authority relations between teachers and their superiors. Effects on teachers were the chief outcome examined, usually gauged in terms of teachers' preferences or satisfactions. In fact, two major reviews spanning the era refer to teacher satisfaction as a research 'preoccupation' (Charters, 1963; Lortie, 1973). The more recent of these reviews continues with an appeal: "Perhaps future research can relate such sentiments [teacher satisfaction] to instructional outcomes, thus linking modes of authority to student learning."* Though subsequent research has been stimulated as much or more by other concerns, efforts to identify school determinants of student achievement have indeed expanded, and several studies have reported that teacher morale is positively associated with student achievement.

Among these earlier studies of teacher satisfaction (in which authority relations among staff were examined as an antecedent) is one that indexed satisfaction by the number of transfers teachers sought. This study reported that teachers tended to seek transfers out of schools attended by students from lower-class homes, and into schools attended by upper-class children (Becker, 1953). We cite this study despite the fact that students' social class origin is not the dependent variable in which we are interested, and in fact, by its nature, can be only an antecedent of teacher satisfaction. Since other studies have shown that differences in students' socio-economic status (SES) largely account for differences in student achievement, and since this study (of teacher transfer requests) suggests that SES also affects teacher satisfaction, then a careful assessment of the preponderant causal direction between student achievement and teacher morale requires considering three possible relationships that

---

*Lortie, D.C. 1973, p. 483.

include SES as a third factor: (1) the effect of SES on achievement is mediated by teacher morale; (2) the effect of SES on teacher morale is mediated by student achievement; (3) SES affects both teacher morale and student achievement, with no mediating effect of one on the other.

A study of Atlanta, Georgia, high schools examined rates of actual teacher transfers and resignations, rather than requests, in an effort to assess the schools' contribution to students' scores on standardized achievement tests (Burkhead et al, 1967). To eliminate the influence of IQ, 10th-grade median scores on the achievement tests were predicted by 8th-grade IQ scores, and the residual values were used as the dependent variable in a regression analysis. Predictor variables included family income and seven status type variables, e.g., buidling age, library expenditure, faculty salary, etc. Only teacher turnover (mobility) was statistically significant: a high rate of teacher mobility was associated with lower test scores.* However, teacher turnover rates were determined from changes in teacher rosters taking place a full year after the students were tested, which suggests that student achievement may be viewed as an input and teacher turnover as the outcome.** Although the authors conclude that the major determinants of academic performance in Atlanta high schools were factors external to the school, such as family income and housing, they do not discuss how these factors might also affect faculty morale and turnover rates.

A study of compensatory programs in elementary schools in Michigan provides more recent and some of the strongest evidence that teacher satisfaction and student achievement gains are related (Michigan, 1976). Results of two study years are reported. Teachers were asked to recall and describe the faculty's morale at the beginning of each fall term, from which a school measure of morale was constructed. Although prior-year gains were the basis for selecting schools, study-year gains (the outcome) were evidently sufficiently similar such that no

*The authors state that the effects of family income were 'parcelled out' (sic) by IQ.

**Students were tested in the fall of 1960; changes between the 1960-61 and 1961-62 teacher rosters determined teacher turnover.

reclassification of the schools and no special study of stability were undertaken.* A positive association between teacher morale and study-year achievement gains was found in four separate analyses. Reviewing these findings, Education Daily (June 25, 1976) addressed the issue of causality, commenting, "...it's still impossible to tell whether the children are reading better because their teachers are happier or the other way around...". However, the Michigan researchers in fact present a path model in which morale is a direct antecedent of achievement. This model conforms to the sequence of their data collection; that is, teacher morale (an input) was assessed prior to the achievement gains (the outcome), justifying the hypothesis that morale affects achievement (more than vice versa). On the other hand, given that prior-year gains were used to select successful and unsuccessful schools and that classifications were not changed, the competing hypothesis that student achievement affects teacher morale, is also supported. Unfortunately, this study did not examine student background, except to compare the median family income of the districts in which successful and unsuccessful schools were located. Such differences between districts were not statistically significant.

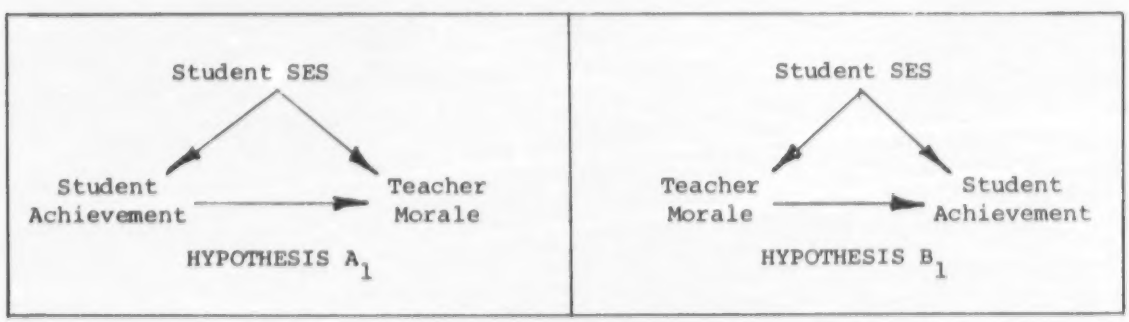
Another recent study, the California School Effectiveness Study (1977), found a positive relationship between teacher satisfaction and academic achievement, but in contrast to the Michigan study, the relationship appeared less strong than the several other variables reportedly associated with achievement. This study considered students' socio-economic status by selecting and matching successful and nonsuccessful schools on this variable. School success was defined by the achievement level of 6th-grade students that was over and above that which was predicted by SES and three other non-school factors. A degree of outcome stability was ensured by basing the classifications on test scores from two successive years. However, these two years preceded the year in which the independent variable data were collected. On the basis of the order in which the data were collected, this study supports the hypothesis 'student achievement affects teacher morale' but, given the apparent stability of achievement in these schools, either of the two major competing hypotheses is tenable. However,

*From private correspondence.

by controlling for student SES, this study contributes substantially to the question of its impact on teacher morale and achievement. From this study one would conclude that there is some effect of achievement on morale, or of morale on achievement, that is not due to students' social origins.

In summary, since most of these studies used a measure of achievement that preceded the assessment of teacher morale, the evidence from the literature favors the hypothesis that student achievement affects teacher morale more than the other way around. However, since the schools in these studies appear to have been consistently successful or nonsuccessful over several years, the competing hypothesis is also tenable. Although the studies reviewed here provide evidence that the relationship between morale and achievement is not completely spurious, it also appears that students' socio-economic status affects both teacher morale and student achievement. This suggests that a revision of the two competing hypotheses to include a third variable, student SES, may be appropriate.* These are diagrammed in Figure 28 which will be useful in the discussion of the ESAA in-depth studies that follows.

Figure 28. The Two Major Competing Hypotheses Explaining A Relationship Between Student Achievement and Teacher Morale, Modified to Include Student Socio-Economic Status



*It should not be inferred that SES is the only other relevant variable: it is included because of its involvement in the potential spuriousness of a relationship between morale and achievement.

4. COMPARISON OF RELATIONSHIPS AMONG TEACHER MORALE AND ALTERNATIVE OUTCOME MEASURES IN THE ESAA IN-DEPTH STUDIES

The outcome measure originally used in the two in-depth studies was based on achievement gains that took place during the study year (see Figure 26, Outcome B). Since the schools were selected for study on the basis of achievement gains in the previous year, an 'outcome' based on these gains was also available (see Figure 26, Outcome A). Data for a third 'outcome', prior achievement level, was available in the form of study-year pretest scores (see Figure 26, Test B-1). Each of these 'outcomes' was tested for an association with teacher satisfaction. The assessment of causality was based on a comparison of these associations.

Before presenting the results of this assessment, we will review relevant aspects of the two ESAA in-depth studies: sample selection; data collection and sources; the criteria used to define the outcomes; and the measures of teacher satisfaction. Although very similar, the two in-depth studies are not exactly the same in all respects. For additional details, see appended tables.

a. Overview of Relevant In-Depth Study Procedure

The 1974-75 study sample consisted of 24 elementary schools drawn from the 101 schools that were participating in the second year of the ESAA evaluation. The 1975-76 sample consisted of 26 schools drawn from the 78 that were participating in the third and final year of the evaluation. Schools were purposefully drawn from the extremes of the previous year's achievement gain distribution in order to obtain distinct comparison groups.

Each of the in-depth study schools was visited by a two-person interviewer/observer team for two weeks during the winter of the study year. During their visit, the teams administered questionnaires and interviews to principals and teachers, and used structured protocols to record observations. The analysis of teacher satisfaction relied primarily on interviews with teachers. Student socio-economic and achievement data were collected in connection with the evaluation of all ESAA study schools.

b. Outcome Measures for the Present Analysis

(1) Selection-Year Gains in Student Achievement

In the original in-depth studies, ESAA schools were designated successful or nonsuccessful according to their achievement gain in the year prior to the study year, in order to select the in-depth sample. In the present analyses, this designation is used as an 'outcome'; prior year success/nonsuccess corresponds to Outcome A in Figure 26. By this definition fifteen of the 1974-75 in-depth schools were successful and nine were nonsuccessful; fourteen 1975-76 schools were successful and twelve nonsuccessful.*

(2) Study-Year Gains in Student Achievement

In the original studies, the outcome was defined in terms of the number of grades that improved in national percentile standing between the fall and spring of the study year. In the present analysis, study-year success/nonsuccess is also used as an outcome and corresponds to Outcome B in Figure 26.

As in selecting the schools, the objective was to identify schools with consistent patterns of high or low achievement gains. ESAA sample students in three grades (3, 4, and 5) were retested in the two basic skills (reading and math). Schools were classified successful if at least two grades gained in national percentile standing in one of the two basic skills. In an effort to produce more distinct comparison groups, classification in 1975-76 also required that

---

*Prior gain of the 1974-75 study schools was in terms of posttest scores adjusted for pretest; prior gain of the 1975-76 study schools was in terms of the number of grades improving in national percentile standing.

In the first study, schools that ranked in the upper 40 percent of the 1975-74 achievement gain distribution were defined as successful; those in the lower 40 percent were nonsuccessful.

In the second study, successful schools were those in which two of three grades improved in national percentile standing in both reading and math; nonsuccessful schools were ones in which no more than one grade improved in each skill area. Since only 10 of the 78 third-year ESAA schools met the cross-skill success criterion, three other schools were included in the successful category. In these three schools, all three grades improved in one skill area and one grade improved in the other area.

at least one grade in the other subject area should show gain.* Nonsuccessful schools were defined as 'all others' in 1974-75; but in 1975-76 the classification in this group was also more stringent, requiring that no more than one grade improved in either area. Application of these criteria identified 14 successful and 10 nonsuccessful schools in 1974-75, and 9 successful and 13 nonsuccessful schools in 1975-76. The three 1975-76 schools that met neither the success nor the nonsuccess criteria were excluded from relational analyses with this outcome,** and a fourth school was dropped because data on only one grade were available.

(3) Achievement Level

The achievement level measure that was used as an 'outcome' in this analysis was a school average of the study-year pretest scores. In Figure 26, this is shown as 'Test B-1'. To obtain the school measure, test scores were standardized by grade on the sample, separately for reading and math, and then averaged for the school.

(4) Teacher Satisfaction Measure

After raising the question of school morale in a general way, interviewers presented teachers with the following statements:

- Teaching in the school is rewarding.
- The school has a good reputation to outsiders; teachers want to come to the school.
- Teachers and staff are a team working together to solve problems.

---

*In five 1975-76 schools, gain scores were available for only two grades. Classification of three of these schools was unambiguous, since all grades either gained or dropped in both subjects. Classification of the other two schools was judgmental (see p. II-7 in An In-Depth Study of Emergency School Aid Act (ESAA) Schools: 1975-1976 by Wellisch et al, which is available from Dr. David Shoemaker, USOE).

**In the present analysis they are included in the analyses with other 'outcomes'.

- Most teachers stay in the school because they have no choice.
- The problems are overwhelming; little can be accomplished.

A school-level scale measure of teacher satisfaction was constructed from their responses to these five statements. In the second study, the responses that interviewers requested and the manner in which the responses were aggregated were modified in an effort to increase measurement sensitivity. Consequently, scale scores for the two studies are not fully comparable. (The interview, aggregation, and scale construction procedures are described in the appended tables.) Despite this, as shown by the distribution in Table 95, the scores for both years appear to cluster fairly densely within a relatively narrow range of possible scores. We interpret 1974-75 scores of 5 as representing roughly the same expression of satisfaction as 1975-76 scores of 2.25 and above.

Table 95. Distribution of Teacher Satisfaction Scale Scores

Interpretation of Scores	1974-75 ESAA In-Depth Schools		1975-76 ESAA In-Depth Schools	
	Score[a] (Median=4)	Number of Schools (N=24)	Score[b] (Median=2.23)	Number of Schools[c] (N=25)
High Satisfaction	5	11	2.75 to 3	1
.	4	8	2.25 to 2.74	11
.	3	2	1.75 to 2.24	11
.	2	3	1.25 to 1.74	2
Low Satisfaction	0-1	0	0 to 1.24	0

[a]Scale scores for the 1974-75 study ranged from a possible 0 to 5, integer values only; actual scores ranged from 2 to 5.

[b]Scale scores for the 1975-76 study ranged from a possible 0 to 3, decimal values included; actual scores ranged from 1.36 to 2.83.

[c]The teacher satisfaction scores of the three 1975-76 schools that were excluded from analyses of study-year gains were 1.96, 2.47 and 2.63.

c. Results

In both the 1974-75 and the 1975-76 ESAA in-depth study schools the relationship between teacher satisfaction and student achievement was virtually the same, regardless of which achievement measure was used to test the association. Teacher satisfaction and student achievement gain were unrelated, whether school success in raising student achievement was measured during the prior year (Outcome A) or the study year (Outcome B). (See Tables 97 and 98.) However, achievement level was positively and significantly related to teacher satisfaction: the higher the school average pretest scores in reading and math (Test B-1), the greater the level of morale among teachers at the school. The relationship was comparably moderate both study years: $r = .35$ (in 1974-75) and $r = .40$ (in 1975-76), and significant at the .05 level. When examined separately for the two basic skill areas, the relationship was slightly stronger for reading than for math.

Since achievement level was assessed at the beginning of the study-year, several months before teacher satisfaction was assessed, these results tend to support the hypothesis that student achievement affects teacher morale (Hypothesis A) more than vice versa (Hypothesis B). However, we did assess achievement level at the end of the year as well and, not surprisingly, since pretest scores were the best predictors of posttest scores, the correlations of this measure of achievement level with teacher satisfaction are of nearly the same magnitude as those of the pretest measure ($r = .36$, 1974-75; and $r = .46$, 1975-76).*

More compelling evidence of a preponderant causality--that student achievement affects teacher morale more than vice versa--is the fact that the satisfaction

---

*The slightly higher correlations with posttest performance, (.35/.36 and .40/.46) suggest two possibilities. One is that teacher morale does affect student achievement (Hypothesis B), however marginally. A second possibility concerns the information teachers possess about achievement level, and whether at the time their morale is assessed it is more accurately reflected by posttest than pretest performance. Earlier we stated that we would assume 'teachers are informed of test results' to be an invariant intervening variable. Here it is pertinent to consider that in addition to test results, teachers certainly possess other information about their students' performance by midyear, which is likely to be evident in posttest scores.

teachers experienced correlated with the level at which their students performed rather than on the gains they achieved. Achievement level, does, of course, consist in part of previous gains, some of which are presumably a consequence of the school attended, and therefore may be partly determined by the levels of morale that existed earlier. But, given that teacher satisfaction and student achievement gains in a single year were unrelated, the hypothesis that teacher satisfaction affects achievement level (by affecting gain) is not supported.

One final step in analyzing these data remained. Because achievement level rather than gain was associated with teacher satisfaction, and because other studies have shown that achievement and socio-economic status (SES) covary strongly, it seemed important to determine whether the relationship might be spurious; specifically, we wished to know whether the 'true determinant' of teacher morale might be student SES.

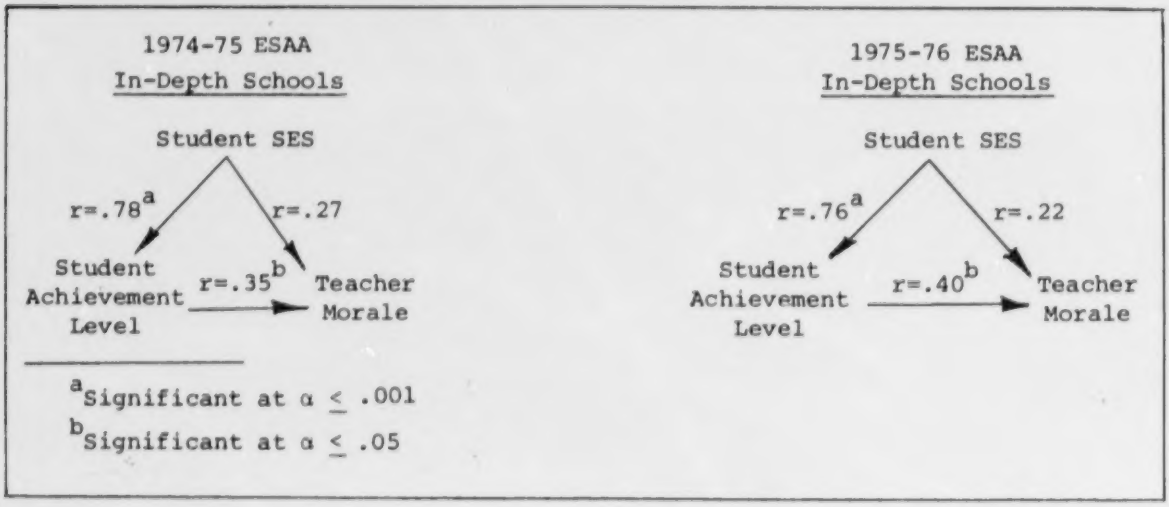
We therefore first tested the ESAA in-depth data for the other two relationships posted in Figure 28--student SES with student achievement (study year pretest scores), and student SES with teacher satisfaction.* The results are displayed in Figure 29, together with Hypothesis A_1 (student achievement affects teacher morale, both of which are affected by student SES) in the form shown earlier in Figure 28, since that is now the favored hypothesis.

The first thing to note about these correlations is that the relationship between SES and achievement that has been reported by other studies was again confirmed: the higher the students' socio-economic status, the greater their academic achievement. The relationship was very strong and highly siginficant ($\alpha \leq .001$) in both samples. The second correlation of interest is that of student SES with teacher morale. Although the association was positive--the higher the SES, the better the morale--it was low and not statistically significant. Together, these correlations suggest that teachers are more strongly influenced

*SES was based on students' responses to a checklist of 15 items present in their homes, five of which may be classified as 'reading related' (e.g., dictionary), and ten as 'luxury', e.g., dishwasher).

by how their students perform academically than by the kinds of homes they come from, even though this is highly correlated with their performance.

Figure 29. Zero-Order Correlations Among Variables in Hypothesis A



However, because it appeared that teachers were at least marginally affected by student background, it seemed useful to attempt to establish whether and to what extent the effect of student achievement on teacher morale could be explained by the fact that higher-performing students came from more advantaged backgrounds. Therefore, we tested the relationship between student achievement level and teacher satisfaction while controlling statistically for SES. The results indicate that while the effect of student achievement level is not as great when SES is controlled for, it is not much less than when SES is ignored. Partial correlations were somewhat weaker than the simple zero order correlations, but not greatly so; they dropped from .35 to .27 (NS) for the 1974-75 schools and from .40 to .36 ($\alpha \leq .05$) for the 1975-76 schools.

In brief, the evidence from the two ESAA in-depth studies is that student achievement affects teacher morale more than morale affects achievement; that it is students' level of performance rather than their progress that is critical; and that the effect of student achievement on teacher morale could only in small

part be accounted for by the fact that higher achieving students tended to be socio-economically more advantaged.

More generally, these results illustrate that findings should be interpreted with care. When attempting to assess the causal direction between student achievement and program or input variables, the time sequences involved should be considered--whether the inputs are measured before or after an achievement gain is measured, or whether it is achievement level (which largely represents prior gain) that is the outcome.

Table 96. Construction of the Teacher Satisfaction Measure

INTERVIEW PROCEDURES

1974-75 - Each teacher was asked if she/he agreed or disagreed with each of the statements (a through e, see text).

1975-76 - Each teacher was asked to judge how other teachers at the school felt; that is, whether (1) all or most, (2) some, (3) few, or (4) no teachers would agree with each statement (a through e, see text.

AGGREGATION PROCEDURES

1974-75 - For statements a, b, and c, a school-level code of 1 indicated 'satisfaction' if 2/3 or more of the teachers agreed; and for statements d and e, if 2/3 or more disagreed.

1975-76 - For statements a, b, and c, response options were coded: all or most = 3, some = 2, few = 1 and no = 0. For statements d and e, response options were coded all or most = 0, some = 1, few = 2, and no = 3. A school-level measure of satisfaction was obtained by averaging the teachers' responses.

SCALE CONSTRUCTION

1974-75 - The scale measure of satisfaction was the number of statements coded 1; that is, the number of statements for which responses of 2/3 or more of the teachers indicated satisfaction.

1975-76 - The scale measure of satisfaction was the average of the scores to the five statements; that is, the grand average at a school of all responding teachers to all five statements.

Table 97. Comparison of Mean Teacher Satisfaction Scale Scores in Successful and Nonsuccessful Schools in Prior Year Achievement Gains (Outcome A)

1974-75 ESAA In-Depth Schools	Teacher Satisfaction Mean Scale Score
Successful (N=15)	4.3
Nonsuccessful (N=9)	3.8
1975-76 ESAA In-Depth Schools	
Successful (N=13)	2.1
Nonsuccessful (N=13)	2.3

Table 98. Comparison of Mean Teacher Satisfaction Scale Scores in Successful and Nonsuccessful Schools in Study Year Achievement Gains (Outcome B)

1974-75 ESAA In-Depth Schools	Teacher Satisfaction Mean Scale Score
Successful (N=14)	4.2
Nonsuccessful (N=10)	4.0
1975-76 ESAA In-Depth Schools	
Successful (N=9)	2.1
Nonsuccessful (N=13)	2.2

CORPORATE REFERENCES

Combined ESAA Pilot and Basic Program Evaluations: Final Study Design. System Development Corporation, TM-5236/002/00, January 1974.

Combined ESAA Pilot and Basic Program Evaluations: Final Study Design, Volume II. System Development Corporation, TM-5236/007/00, August 1975.

Combined ESAA Pilot and Basic Program Evaluations: Final Study Design, Volume III. System Development Corporation, TM-5236/015/00, February 1977.

Coulson, J. E., Ozenne, D. G., Van Gelder, N. C., Inuzuka, D., Bradford, C., and Doherty, W. J. The First Year of Emergency School Aid Act (ESAA) Implentation. System Development Corporation, TM-5236/008/00, September 1975.

Coulson, J. E., Ozenne, D. G., Bradford, C., Doherty, W. J., Duck, G. A., Hemenway, J. A., and Van Gelder, N. C. The Second Year of Emergency School Aid Act (ESAA) Implementation. System Development Corporation, TM-5236/009/00, July 1976.

Coulson, J. E., Ozenne, D. G., Hanes, S. D., Bradford, C., Doherty, W. J., Duck, G. A., and Hemenway, J. A. The Third Year of Emergency School Aid Act (ESAA) Implementation. System Development Corporation, TM-5236/014/00, March 1977.

Coulson, J. E. National Evaluation of the Emergency School Aid Act (ESAA): Summary of the Second-Year Studies. System Development Corporation, TM-5236/011/00, July 1976

Coulson, J. E. Overview of the National Evaluation of the Emergency School Aid Act. System Development Corporation, TM-5236/022/00, July 1977

Cromer, F. E. The Distribution of Achievement Scores in a Disadvantaged Population: Data From a National Sample of Students in Districts Receiving Emergency School Aid Act (ESAA) Funds. System Development Corporation, TM-5236/026/00, In Press, March 1978.

Doherty, W. J. The Restandardization Study in Combined ESAA Pilot and Basic Program Evaluations: ESAA Study Design Volume III. System Development Corporation, TM-5236/009/00, 1976.

MacQueen, A. H. Emergency School Aid Act (ESAA) Evaluation: Overview of Findings From Supplemental Analyses. System Development Corporation, TM-5236/027/00, In Press, March 1978.

Ozenne, D. G., Van Gelder, N. C., and Cohen, A. J. Emergency School Aid Act (ESAA) National Evaluation: Achievement Test Restandardization, TM-5236/006,00, November 1974.

Wellisch, J. B., Marcus, A. C., MacQueen, A. H., and Duck, G. A. An In-Depth Study of Emergency School Aid Act (ESAA) Schools: 1974-1975. System Development Corporation, TM-5236/010/00, July 1976.

Wellisch, J. B., Carriere, R. A., MacQueen, A. H., and Duck, G. A. An In-Depth Study of Emergency School Aid Act (ESAA) Schools: 1975-1976. System Development Corporation, TM-5236/013/00, March 1977.

Wellisch, J. B., MacQueen, A. H., Carriere, R. A., Duck, G. A.. Some Policy-Relevant Findings in the ESAA In-Depth Studies: Organizational Climate and Equality of Educational Opportunity. System Development Corporation, TM-5236/023/00, August 1977

GENERAL REFERENCES

Angoff, W. H. Scales, Norms, and Equivalent Scores, in Thorndike, R. (Ed) Educational Measurement. Second Edition, Washington, D.C.: American Council on Education, 1971.

Barker, P. and Pelavin, S. Conserving Scores and Scale Transformations in Standardized Achievement Tests, Their Accuracy and Dependability for Individual and Aggregation: The Case of MAT70. Rand Corporation, Santa Monica, Working Note, WN-9161-NIE, 1975.

Barr, J. A., Goodnight, J. H., Sall, J. P., and Helwig, J. T. A User's Guide to SAS 76. Raleigh: SAS Institute Inc., 1976.

Becker, H. S. The Teacher in the Authority System of the Public School. Journal of Educational Sociology. 1953, 27, 128-141.

Bryk, A. S. and Weisberg, H. I. Use of the Nonequivalent Control Group Design When Subjects are Growing. Psychological Bulletin, 84, 1977, 950-962.

Burkhead, J., Fox, T. G., Holland, J. W., Input and Output in Large City High Schools. Syracuse University Press, Syracuse, 1967.

California State Department of Education. California School Effectiveness Study, The First Year: 1974-75. Sacramento, California.

Campbell, D. T. and Stanley, J. C., Experimental and Quasi-Experimental Designs for Research. Skokie, Illinois: Rand McNally, 1966.

Charters, W. W., Jr. The Social Background of Teaching, in Gage, N. L. (Ed) Handbook of Research on Teaching. Chicago: Rand McNally, 1963, 474-497.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. A., Weinfield, F. D., and York, R. L., Equality of Educational Opportunity. Washington, D.C.: U.S. Office of Education, 1966.

Coulson, J. E. National Evaluation of the Emergency School Aid Act (ESAA): A Review of Methodological Issues. Journal of Educational Statistics. (in press)

Cureton, E. E. and Tukey, J. W. Smoothing Frequency Distributions, Equating Tests, and Preparing Norms. American Psychologist, 1951, 6, 404.

Education Daily, Michigan Learns How to Buy Better Reading Achievement. Vol. 9, 1976, No. 124, June 25, 1976.

Fernandez, C., Espinosa, R., and Dornbusch, S. M. Factors Perpetuating the Low Academic Status of Chicano High School Students. Stanford Center for Research and Development on Teaching, Palo Alto, 1975.

Haggart, S. Program Cost Analysis in Educational Planning, Santa Monica: Rand Corporation, 1971.

Harman, H. H., Modern Factor Analysis. Chicago: The University of Chicago Press, 1967.

Hilton, T. L. and Patrick, C. Cross-Sectional Versus Longitudinal Data: An Empirical Comparison of Mean Differences in Academic Growth. Journal of Educational Measurement, Vol. 7, No. 1 (spring, 1970), 15-24.

Johnson, S. C., Hierarchical Clustering Schemes, Psychometrika, XXXII (1967), 241-254.

Kaskowitz, D. H. and Normwood, C. R. A Study of the Norm-Referenced Procedure for Evaluating Project Effectiveness as Applied in the Evaluation of Project Information Packages. Stanford Research Institute, Research Memorandom, Menlo Park, California, 1977.

Long, J. Schaffran, J. and Kellog, T. Effects of Out-of-Level Survey Testing on Reading Achievement Scores of Title I ESEA Students. Journal of Educational Measurement, 1977, 14, 203-213.

Lord, F. M. Estimation of Latent Ability and Item Parameters When There are Omitted Responses. Psychometrika, 1974, 39, 247-264.

Lortie, Dan C. Observations on Teaching as Work in Second Handbook of Research in Teaching, Travers, W. (Ed), Chicago: Rand McNally, 1973, 715-813.

Michigan Department of Education. Report of the 1974-75 Michigan Cost-Effectiveness Study. Washington, D.C.: Capital Publications, 1976.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., and Bent, D. H. SPSS: Statistical Package for the Social Sciences. New York: McGraw-Hill Book Company, 1975.

Pelavin, S. and Barker, P. A Study of the Generalizability of the Results of a Standardized Achievement Test. Paper presented at the American Educational Research Association Meeting, San Francisco, California, April 1976.

Pelavin, S. and David, J. Evaluating Long Term Achievement: An Analysis of Longitudinal Data from Compensatory Education Programs. Stanford Research Institute, EPRC 4537-15, 1977.

Slinde, J. A. and Linn, R. L. Vertically Equated Tests: Fact or Phantom? Journal of Educational Measurement, 1977, 14, 23-32.

Thurstone, L. L. A Method of Scaling Psychological and Educational Tests. Journal of Educational Psychology, 1925, 16, 444-451.

Thurstone, L. L. Primary Mental Abilities. Psychometric Monographs No. 1 Chicago: University of Chicago, 1933.

Tiegs, E. W. and Clark, W. W. Bulletin of Technical Data No. 1, California Achievement Tests. 1970 Edition, Monterey, California: CTB/McGraw-Hill, 1970.

Wallen, Norman E. and Travers, Roberts, M. W. Analysis and Investigation of Teaching Methods in Handbook of Research on Teaching, Gage, N. W. (Ed), Chicago: Rand McNally and Company, 1963.